

**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE SISTEMAS E
INFORMAÇÃO**

APOLO TAKEHI ARAI BATISTA

**DETECÇÃO DE FRAUDE EM SOCIAL COMMERCE: UMA ABORDAGEM
BASEADA NA COMBINAÇÃO DE INFORMAÇÕES ESTRUTURADAS E
IMAGENS**

**RIO DE JANEIRO
2020**

APOLO TAKEHI ARAI BATISTA

DETECÇÃO DE FRAUDE EM SOCIAL COMMERCE: UMA ABORDAGEM
BASEADA NA COMBINAÇÃO DE INFORMAÇÕES ESTRUTURADAS E
IMAGENS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Informação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Sistemas e Informação .

Orientador(es): Ronaldo Goldschmidt, D.Sc.
Karla Figueiredo, D.Sc.

Rio de Janeiro

2020

©2020

INSTITUTO MILITAR DE ENGENHARIA
Praça General Tibúrcio, 80 – Praia Vermelha
Rio de Janeiro – RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmар ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Takehi Arai Batista, Apolo.

Detecção de Fraude em Social Commerce: uma abordagem baseada na combinação de informações estruturadas e imagens / Apolo Takehi Arai Batista.

– Rio de Janeiro, 2020.

154 f.

Orientador(es): Ronaldo Goldschmidt e Karla Figueiredo.

Dissertação (mestrado) – Instituto Militar de Engenharia, Engenharia de Sistemas e Informação, 2020.

1. detecção de fraude. 2. social commerce. 3. s-commerce. 4. e-commerce.
5. comércio eletrônico. 6. aprendizado de máquina. 7. avaliação de imagens.
8. imagem. i. Goldschmidt, Ronaldo (orient.) ii. Figueiredo, Karla (orient.) iii.
Título

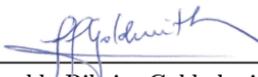
APOLO TAKEHI ARAI BATISTA

**Detecção de Fraude em Social Commerce: uma
abordagem baseada na combinação de informações
estruturadas e imagens**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Informação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Sistemas e Informação .

Orientador(es): Ronaldo Goldschmidt e Karla Figueiredo.

Aprovado em Rio de Janeiro, 09 de dezembro de 2020, pela seguinte banca examinadora:



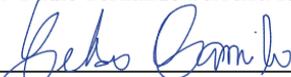
Prof. **Ronaldo Ribeiro Goldschmidt** - D.Sc. do IME



Prof. **Karla Tereza Figueiredo Leite** - D.Sc. da UERJ



Prof. **Paulo Fernando Ferreira Rosa** - D.Sc. do IME



Prof. **Celso Gonçalves Camilo Junior** - D.Sc. da UFG

Rio de Janeiro
2020

*Este trabalho é dedicado aos meus pais, que
sempre me inspiraram a seguir estudando,
e aqueles cuja pesquisa possa
ajudar de alguma forma.*

AGRADECIMENTOS

Os agradecimentos principais são direcionados à Karla Figueiredo e ao Ronaldo Goldschmidt sem os quais eu não teria concluído este curso e a todos aqueles que contribuíram para minha formação acadêmica e pessoal no IME, como meus colegas e professores. Em especial gostaria de agradecer à professora Justel por ter feito uma matéria tão rica e interessante, amei o desafio das suas provas! Aos professores Alex Garcia e Marcos Veloso que me receberam como aluno fazendo apenas a matéria deles e me incentivaram a me matricular de fato no mestrado. E um obrigado muito especial ao Vicente, Lamoglia e Duarte que sempre me acolheram com quaisquer problemas me permitindo sempre resolver as situações rápida e prontamente.

*“The fraudster’s greatest liability is
the certainty that the fraud is
too clever to be detected.”
(Louis Freeh(1), ex-diretor do FBI)*

RESUMO

As necessidades de mercado demandaram mudanças nos *e-commerce* e nas redes sociais, resultando no surgimento do *social commerce*. Somado a isto, é observado o contínuo crescimento das transações comerciais em *Social Commerce*. Este ambiente é atraente para usuários bem e mal intencionados, os últimos causam danos financeiros e psicológicos as suas vítimas por meio das fraudes online. Como a volumetria de casos é alta e a ocorrência de fraude é baixa, o processo de detecção manual não é escalável e é ineficiente, muito recurso para pouca detecção. Por outro lado, as soluções existentes para a detecção automática de fraude em *Social Commerce* baseiam-se em informações estruturadas extraídas dos anúncios (i.e.: preço, se é novo ou usado, qual a cor, etc.). Entretanto, tais soluções ignoram potenciais indícios de fraude nas imagens que complementam as informações sobre o objeto comercializado. Assim, o presente trabalho propõe uma investigação da utilização de *Deep Learning* para combinação de diferentes algoritmos unindo base de dados estruturadas e imagens. Desse modo, foi feita uma proposta de um método chamado de DFSC (Detecção de Fraude em *Social Commerce*), que permitiu a análise de anúncios considerando suas respectivas imagens e informações estruturadas. Para tais combinações foram estudados a utilização de modelos de dados estruturados isoladamente, de imagens isoladamente, e duas formas de combinação dos modelos seja considerando a saída dos modelos de imagem como um dado enriquecido para um modelo de dados estruturado, seja por meio da composição de modelos de dados estruturados e de imagens sujeitos a um novo classificador para avalia-los. Tais experimentações se mostraram promissoras, apresentando resultados que evidenciam que a consideração da imagem impacta de modo significativo a qualidade de detecção de fraude em *Social Commerce*, com ganhos de até 20% em F-Score.

Palavras-chave: detecção de fraude. social commerce. s-commerce. e-commerce. comércio eletrônico. aprendizado de máquina. avaliação de imagens. imagem.

ABSTRACT

The Social commerce has risen due to changes in either e-commerce or social networks. On top of that, the number of online ads and transactions in Social Commerce has grown. This environment is attractive to either good users and bad users. The bad users cause harm to their victim by making them lose money or psychological damage. Since the volume of transactions is high and the fraud occurrence is low, the manual detection is not scalable and highly inefficient, too much resource wasted for low detection. The existing solutions for automatic fraud detection in Social Commerce, e-commerce with social networks, are based on structured information available in ads such as price, product type, brand, new/used, among others. However, such solutions ignore possible fraud signs from the ads' images that exhibit the product sold. Therefore, this article aims to evaluate if combining structured information and images available in the ads provides more effective models than the ones that consider only structured information. The proposed method uses deep learning to evaluate the images. It has been developed a method called Fraud Detection at Social Commerce (FDSC) which has enabled the combined evaluation of ads' images and structured information. Experimental evidence shows an incremental opportunity of 20% in F-score by the adoption of this method.

Keywords: fraud detection. social commerce. s-commerce. e-commerce. machine learning. image evaluation. image.

LISTA DE ILUSTRAÇÕES

Figura 1 – Acompanhamento do Número de Reclamações ao longo do tempo (2) . . .	23
Figura 2 – Perda Financeira registrada em Dólar(USD) ao longo do tempo (2) . . .	23
Figura 3 – Representação de locais onde ocorrem e-commerce	28
Figura 4 – Redes Sociais + E-commerce = Representação Visual do Social Commerce	30
Figura 5 – Detalhamento de Alvos de Fraude (3)	31
Figura 6 – Detalhamento de Fraude Contra o Indivíduo (3)	31
Figura 7 – Detalhamento de Fraude Contra Organizações (3)	32
Figura 8 – Ilustração aprendizado supervisionado e não supervisionado	33
Figura 9 – Proposta de Modelo de Dados Conceitual	38
Figura 10 – DFSC: Visão Geral das Alternativas de Aplicação	48
Figura 11 – Visão Geral das Etapas Comuns à Utilização do DFSC	49
Figura 12 – DFSC(Est): Visão Geral das Etapas	50
Figura 13 – DFSC(Img): Visão Geral das Etapas	51
Figura 14 – DFSC(EST+Img): Visão Geral das Etapas	53
Figura 15 – DFSC(EST,Img): Visão Geral das Etapas	55
Figura 16 – Exemplo 01 apresentado no site (4)	60
Figura 17 – Exemplo 02 apresentado no site (4)	61
Figura 18 – Modelos Img - Arquitetura <i>CNN TransferLearning</i> (ImageNet adaptada)	69
Figura 19 – Função de Perda da Melhor Arquitetura de Imagem	70
Figura 20 – Aplicação da melhor arquitetura obtida por DFSC(Img) ao conjunto DI_{val}	70
Figura 21 – Visão Geral DFSC(Est,Img)	79
Figura 22 – DFSC(Est) - Usando GNB como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	98
Figura 23 – DFSC(Est,Img) - Usando GNB como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	99
Figura 24 – DFSC(Est) - Usando KNN como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	101
Figura 25 – DFSC(Est,Img) - Usando KNN como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	102
Figura 26 – DFSC(Est) - Usando RF como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	103
Figura 27 – DFSC(Est,Img) - Usando RF como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	104
Figura 28 – DFSC(Est) - Usando MLP como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	106

Figura 29 – DFSC(Est,Img) - Usando MLP como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	107
Figura 30 – DFSC(Est) - Usando LGR como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	108
Figura 31 – DFSC(Est,Img) - Usando LGR como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	110
Figura 32 – DFSC(Est) - Usando LDA como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	111
Figura 33 – DFSC(Est,Img) - Usando LDA como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	112
Figura 34 – DFSC(Est) - Usando DCT como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	114
Figura 35 – DFSC(Est,Img) - Usando DCT como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor	115
Figura 36 – DFSC(Est,Img) - Usando GNB como combinador e analisando a inclusão dos dados da imagem	117
Figura 37 – DFSC(Est,Img) - Usando KNN como combinador e analisando a inclusão dos dados da imagem	118
Figura 38 – DFSC(Est,Img) - Usando RF como combinador e analisando a inclusão dos dados da imagem	119
Figura 39 – DFSC(Est,Img) - Usando MLP como combinador e analisando a inclusão dos dados da imagem	120
Figura 40 – DFSC(Est,Img) - Usando LGR como combinador e analisando a inclusão dos dados da imagem	121
Figura 41 – DFSC(Est,Img) - Usando LDA como combinador e analisando a inclusão dos dados da imagem	123
Figura 42 – DFSC(Est,Img) - Usando DCT como combinador e analisando a inclusão dos dados da imagem	124
Figura 43 – Visão Geral DFSC(Est,Img)	127
Figura 44 – DFSC(Est,Img) - Usando GNB como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	131
Figura 45 – DFSC(Est,Img) - Usando GNB como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	132
Figura 46 – DFSC(Est,Img) - Usando KNN como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	133
Figura 47 – DFSC(Est,Img) - Usando KNN como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	134
Figura 48 – DFSC(Est,Img) - Usando RF como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	135

Figura 49 – DFSC(Est,Img) - Usando RF como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	137
Figura 50 – DFSC(Est,Img) - Usando MLP como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	138
Figura 51 – DFSC(Est,Img) - Usando MLP como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	139
Figura 52 – DFSC(Est,Img) - Usando LGR como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	140
Figura 53 – DFSC(Est,Img) - Usando LGR como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	141
Figura 54 – DFSC(Est,Img) - Usando LDA como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	142
Figura 55 – DFSC(Est,Img) - Usando LDA como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	144
Figura 56 – DFSC(Est,Img) - Usando DCT como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem	145
Figura 57 – DFSC(Est,Img) - Usando DCT como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem	146
Figura 58 – Visão Geral DFSC(Est,Img) considerando a partir de 2 modelos	148
Figura 59 – DFSC(Est,Img) - Usando KNN como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	149
Figura 60 – DFSC(Est,Img) - Usando RF como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	150
Figura 61 – DFSC(Est,Img) - Usando MLP como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	151
Figura 62 – DFSC(Est,Img) - Usando LGR como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	152
Figura 63 – DFSC(Est,Img) - Usando LDA como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	153
Figura 64 – DFSC(Est,Img) - Usando DCT como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)	154

LISTA DE QUADROS

LISTA DE TABELAS

Tabela 1 – Quantidade de Trabalhos sobre Fraude desde 1995	36
Tabela 2 – Exemplo de Agrupamento das Características dos Objetos	41
Tabela 3 – Exemplo Agrupamento dos atributos em Indivíduos	41
Tabela 4 – Exemplo Agrupamento dos atributos em Ações	42
Tabela 5 – Exemplo Agrupamento dos atributos em Comunidade	43
Tabela 6 – Autores com informações não explícitas	43
Tabela 7 – Publicações de Fraude por Ano desde 2012	44
Tabela 8 – Visão Resumida dos Trabalhos Relacionados	45
Tabela 9 – Tempo computacional dos experimentos	58
Tabela 10 – Tabela com exemplos de atributos utilizados por <i>e-commerce</i> e <i>social commerce</i>	62
Tabela 11 – Resumo Estatístico da Base de Dados	62
Tabela 12 – Resumo Estatístico dos Conjuntos após a <i>Projeção de Dados</i>	63
Tabela 13 – Algoritmo e Parâmetros	64
Tabela 14 – Melhores algoritmos DFSC(Est)	66
Tabela 15 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para DFSC(Est) selecionados	67
Tabela 16 – Experimentação com alteração na quantidade de camadas	68
Tabela 17 – FDSC(Img) melhor arquitetura e melhor modelo aplicado ao conjunto de Teste D'_{tt}	71
Tabela 18 – Melhores algoritmos DFSC(Est+Img)	73
Tabela 19 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para Melhores algoritmos DFSC(Est+Img)	74
Tabela 20 – Tabela Comparativa de Médias \pm Desvio Padrão ($\mu \pm \sigma$) para Melhores Modelos dos Métodos DFSC(Est) e DFSC(Est+Img) aplicados ao conjunto Teste (DEI'_{tt})	75
Tabela 21 – Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem	76
Tabela 22 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para Melhores algoritmos DFSC(Est,Img)	78
Tabela 23 – Comparação DFSC(Est) e DFSC(Est,Img) de Médias \pm Desvio Padrão ($\mu \pm \sigma$) aplicados ao conjunto Teste (DEI'_{tt})	79
Tabela 24 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Combinadores Avaliados após inclusão da Imagem	81
Tabela 25 – Comparação de Médias \pm Desvio Padrão ($\mu \pm \sigma$) para Algoritmos Combinadores Considerando a Abordagem DFSC(Est,Img) aplicados ao conjunto Teste (D'_{tt})	127

Tabela 26 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem - parte 01 / 02	129
Tabela 27 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem - parte 02 / 02	130

LISTA DE ABREVIATURAS E SIGLAS

DFSC	Detecção de Fraude em Social Commerce
Est	Estruturado
Img	Imagem
TCD	Transações Comerciais Digitais
TIC	Tecnologia de Informação e Comunicação
COVID	COrona VIRus Disease (Doença do Coronavírus)
USD	U.S. dollar (moeda oficial dos Estados Unidos)
KNN	K-Nearest Neighbors
LGR	Regressão Logística
MLP	MultiLayer Perceptron
RF	Random Forest
DCT	Árvores de Decisão
LDA	Análise Discriminante Linear
GNB	Naïve Bayes Gaussiano
OCR	do inglês, <i>Optical character recognition</i> , à tradução livre, reconhecimento de caracteres em imagens
CNN	do inglês, <i>Convolutional Neural Network</i> Redes Neurais Convolucionais ou redes profundas
SVM	Support Vector Machine
CA	Características Atuais
CH	Características Históricas
UP	Usuário-Plataforma
UU	Usuário-Usuário
LGPD	Lei Geral de Proteção aos Dados

e-commerce	do inglês <i>electronic commerce</i> , comércio eletrônico
s-commerce	do inglês <i>social commerce</i> , não existe uma tradução exata, comércio eletrônico de redes sociais à tradução livre
tr	Treino
tt	Teste
val	Validação

LISTA DE SÍMBOLOS

ω	Conjunto das probabilidades associadas a cada imagem do anúncio
Ω	Conjunto de métricas geradas a partir de ω
\in	Pertence
\cup	União
\wedge	E
ϕ	Chaves ou Valores de parâmetros utilizados por um modelo estruturado
μ	Média
\pm	Mais ou menos
σ	Desvio Padrão
\leq	Menor ou igual
\iff	Se somente Se
\times	Multipliação
Σ	Somatório
D	Conjunto de dados do Anúncio
DI	Conjunto das Imagens de um anúncio
DE	Conjunto dos Dados Estruturados de um anúncio
D'	Subconjunto de D
a	anúncio
e_a	atributos estruturados de um anúncio
c_a	rótulo de um anúncio (Fraude ou não Fraude)
DE_{tr}	Conjunto de Dados Estruturados, porção treino
DE_{val}	Conjunto de Dados Estruturados, porção validação
DE_{tt}	Conjunto de Dados Estruturados, porção teste
DI_{tr}	Conjunto de Imagens, porção treino

DI_{val}	Conjunto de Imagens, porção validação
DI_{tt}	Conjunto de Imagens, porção teste
S	Conjunto de algoritmos de classificação
s	algoritmo de classificação
f	métrica de avaliação de desempenho
m_{s_k, f_o}	modelo definido pelo algoritmo de classificação s avaliado conforme a métrica f_o
M	Conjunto de modelos m_{s_k, f_o}
DMI	Conjunto de Dados do Modelo de Imagem: imagens após o processamento por um modelo de imagem que transforma as imagens num atributo estruturado
i_a	Rótulo de classificação de uma imagem que pertence a um anúncio a entre fraude e não fraude
Min	Mínimo
Max	Máximo
Q	Quartis
$\bar{\omega}$	Média do valor ω

SUMÁRIO

1	INTRODUÇÃO	22
1.1	CONTEXTUALIZAÇÃO E MOTIVAÇÃO	22
1.2	PROBLEMA	23
1.3	HIPÓTESE	24
1.4	OBJETIVOS DO TRABALHO	25
1.5	MÉTODO	25
1.6	CONTRIBUIÇÕES ESPERADAS	26
1.7	ORGANIZAÇÃO DO TEXTO	26
2	CONCEITOS BÁSICOS	28
2.1	COMÉRCIO ELETRÔNICO (E-COMMERCE)	28
2.2	DEFINIÇÕES DE FRAUDE	30
2.3	APRENDIZADO DE MÁQUINA	32
2.3.1	SUPERVISIONADO	33
2.3.2	NÃO SUPERVISIONADO	34
2.3.3	TÉCNICAS DE MINERAÇÃO DE DADOS	34
2.3.3.1	CLASSIFICAÇÃO - TÉCNICAS TRADICIONAIS	35
2.3.4	APRENDIZADO PROFUNDO	35
3	TRABALHOS RELACIONADOS	36
3.1	MODELO DE DADOS CONCEITUAL DE <i>SOCIAL COMMERCE</i>	37
3.2	PROPOSTA DE MODELO COMPARATIVO	38
3.2.1	ORGANIZAÇÃO CONFORME TÓPICOS	38
3.2.1.1	CLASSIFICAÇÃO EM FUNÇÃO DO TIPO DE INFORMAÇÃO EXPLORADA	39
3.2.1.2	CLASSIFICAÇÃO EM FUNÇÃO DA ORIGEM DA INFORMAÇÃO	39
3.2.1.2.1	OBJETOS	40
3.2.1.2.2	INDIVÍDUOS	41
3.2.1.2.3	AÇÕES - USUÁRIO/USUÁRIO(UU) OU USUÁRIO-PLATAFORMA(UP)	42
3.2.1.2.4	COMUNIDADE	42
3.3	APLICAÇÃO DO MODELO COMPARATIVO	43
4	MÉTODO PROPOSTO	48
4.1	DFSC(EST)	50
4.2	DFSC(IMG)	51
4.3	DFSC(EST+IMG)	52
4.4	DFSC(EST,IMG)	54

5	EXPERIMENTOS E RESULTADOS	58
5.1	PROTÓTIPO	58
5.2	BASE DE DADOS UTILIZADA NOS EXPERIMENTOS	58
5.3	PRÉ-PROCESSAMENTO DOS DADOS	61
5.4	DFSC(EST)	63
5.4.1	DESCRIÇÃO	64
5.4.1.1	CRITÉRIOS DE DESEMPATE	64
5.4.1.2	MELHORES MODELOS	65
5.4.1.3	MELHOR MODELO	65
5.4.2	RESULTADO	65
5.5	DFSC(IMG)	67
5.5.1	DESCRIÇÃO	68
5.5.1.1	CRITÉRIOS DE DESEMPATE	69
5.5.1.2	MELHORES MODELOS	69
5.5.1.3	MELHOR MODELO	69
5.5.2	RESULTADO	70
5.6	CRITÉRIOS PARA VERIFICAÇÃO DE RELEVÂNCIA ESTATÍSTICA	71
5.7	DFSC(EST+IMG)	71
5.7.1	DESCRIÇÃO	71
5.7.2	RESULTADO	72
5.7.3	ANÁLISE DE RESULTADOS	74
5.7.3.1	AVALIAÇÕES E RELEVÂNCIA ESTATÍSTICA DE CADA UMA DAS MÉTRICAS	75
5.8	DFSC(EST,IMG)	76
5.8.1	DESCRIÇÃO	76
5.8.2	RESULTADO	78
5.8.3	ANÁLISE DE RESULTADOS	78
5.8.3.1	AVALIAÇÕES E RELEVÂNCIA ESTATÍSTICA DE CADA UMA DAS MÉTRICAS	80
6	CONCLUSÃO	82
	REFERÊNCIAS	84
	APÊNDICE A – BIBLIOTECAS USADAS	91
	APÊNDICE B – DETALHAMENTO RELEVÂNCIA ESTATÍSTICA DFSC(EST+IMG) PARA CADA UMA DAS MÉ- TRICAS	98

**APÊNDICE C – DETALHAMENTO RELEVÂNCIA ESTATÍSTICA
DFSC(EST,IMG) PARA CADA UMA DAS MÉ-
TRICAS 117**

APÊNDICE D – ESTUDOS COMPLEMENTARES COM DFSC(EST,IMG)126

D.1	DFSC(EST,IMG) - MELHOR OU MELHORES	126
D.1.1	RESULTADO	126
D.1.1.1	AVALIAÇÕES E RELEVÂNCIA ESTATÍSTICA DA INCLUSÃO DA IMAGEM . . .	128
D.1.2	DETALHAMENTO RELEVÂNCIA ESTATÍSTICA DFSC(EST,IMG) PARA CADA UMA DAS MÉTRICAS	131
D.2	DFSC(EST,IMG) - VARIANDO A QUANTIDADE DE MODELOS	147
D.2.1	RESULTADO	147

1 INTRODUÇÃO

1.1 Contextualização e Motivação

A adoção pelo modo eletrônico de comércio (*e-commerce*) tem aumentado, no Brasil (5) e em outros países, de forma geral, devido a facilidades que envolvem a redução do custo operacional e da burocracia, tanto para sua criação quanto para sua manutenção, além da agilidade no lançamento de lojas *online* (6). Em 2017, por meio de transações comerciais digitais (TCD), as vendas de *e-commerce* alcançaram o valor de US\$ 2.3 trilhões, com perspectivas para dobrar em 2021 (7). O crescimento do *e-commerce* é uma evidência de que o hábito da compra *online* pode ser considerado como estabelecido. O isolamento causado pela pandemia de COVID-19, potencializou o crescimento de consumo semanal desde o 16 de março de 2020 chegando ao crescimento de 387% na semana iniciada em 09 de abril de 2020 comparado à semana anterior (8).

O estudo feito por (5) destaca o crescimento de 1 milhão para 40 milhões de brasileiros (entre 2001 e 2012) além da identificação do Brasil como líder em comércio eletrônico na América Latina em relação à participação de compras em *e-commerce* de 59,1%, o que representa 1% do PIB brasileiro.

Somado a isso, percebe-se o contínuo crescimento das redes sociais, onde o Instagram atingiu a marca de 1 bilhão de usuários ativos (9), sendo 95 milhões de usuários brasileiros (10) e o Facebook detém 2.7 bilhões de usuários ativos (11).

Contudo, as necessidades do mercado demandaram mudanças tanto nos *e-commerce* (oferta de funcionalidades de redes sociais) quanto nas redes sociais (além da conexão entre usuários, negociações e vendas de produto), por meio das TCD, cuja adaptação dos modelos caracterizaram o chamado *social commerce* (12, 13).

Pesquisas feitas no cenário brasileiro em 2019, apontam que 58% dos usuários de internet fizeram uma compra de produto / serviço *online* nos últimos 12 meses, além da estimativa de 28 bilhões de dólares gerados pelo setor em 2018 (14).

Esta nova modalidade de compra e venda, via Internet, somada ao elevado valor movimentado, também abre oportunidades para que usuários mal-intencionados pratiquem diversos crimes *online* (15), o que é confirmado pelo aumento do número de reclamações feitas por vítimas ao FBI envolvendo esta modalidade de crime (16), conforme pode ser observado na Figura 1.



Figura 1 – Acompanhamento do Número de Reclamações ao longo do tempo (2)

Entre os tipos de crimes que ocorrem em *e-commerce*, estão as fraudes *online* que, além de trazer grandes prejuízos de ordem econômico-financeira (17), conforme pode ser observado na Figura 2, para empresas e indivíduos, (18), podem causar sérios danos pessoais a estes últimos.

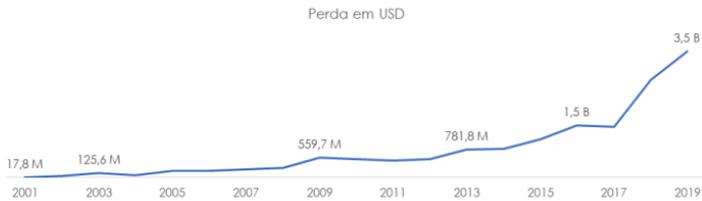


Figura 2 – Perda Financeira registrada em Dólar(USD) ao longo do tempo (2)

Para mitigar a ocorrência desse tipo de crime em ambientes de *e-commerce*, o combate a fraudes em TCD pode ser realizado de forma preventiva, por meio de um processo de monitoramento, que impede a ocorrência da fraude antes de qualquer comprador manifeste interesse no produto, ou por detecção, quando há a falha do modo preventivo, e a transação fraudulenta já teve início (19, 15).

1.2 Problema

O número de usuários, ações e transações que ocorrem no *social commerce* continuam crescendo conforme as empresas divulgam mais funcionalidades aos usuários e se adaptam as suas necessidades. Contudo, até onde foi possível observar com a pesquisa bibliográfica, as pesquisas em *social commerce* costumam ser realizadas para fins de entendimento de estratégias de marketing pois, ora estudam o comportamento do usuário enquanto comprador, ora estudam como o usuário interage com diversas marcas ou uma marca específica(13). Se por um lado há o estudo para entender mais sobre o usuário ideal e

as formas como este interage com o produto e marca, por outro lado, a facilidade em se tornar um vendedor implica num aumento de produtos entendidos como defeituosos, por exemplo, falsificados. É observada a necessidade de se ter soluções cada vez mais escaláveis uma vez que a complexidade do *social commerce* se torna cada vez maior, seja pela oferta de funcionalidades, seja por novas formas de interagir (fotos, áudios, mensagens temporárias, *emojis*, reações) e se a única alternativa for por meio da moderação de conteúdo humana, nota-se um elevado esforço humano para detectar um baixo volume de fraudadores; portanto, ineficiente e não escalável (20).

As principais soluções escaláveis, existentes na área, utilizam modelos de classificação construídos pela aplicação de algoritmos de aprendizado de máquina e podem ser divididas em dois grupos.

As soluções do primeiro grupo se baseiam exclusivamente em informações estruturadas extraídas dos anúncios, tais como preço, tipo de produto, marca, modelo, se é novo ou usado, local físico de venda, entre outras (21, 22, 23). Embora produzam bons resultados, tais soluções ignoram potenciais indícios da fraude nas imagens, que podem complementar o conjunto de informações disponibilizadas sobre o produto comercializado.

No segundo grupo estão os trabalhos que exploram apenas características das imagens (24, 25, 20, 26). Assim, apesar de inovarem na área, tais trabalhos não levaram em conta informações estruturadas associadas aos anúncios. Um dos trabalhos identificados na pesquisa se limitou a explorar apenas atributos presentes nas imagens, e buscando, por meio de agrupamento e modelagem de grafos, identificar perfis de usuários suspeitos (20, 26).

1.3 Hipótese

Diante do problema exposto, a seguinte hipótese foi levantada: *Considerar informações das Imagens e dados estruturados dos anúncios pode levar a modelos de detecção de fraude, em Social Commerce, com desempenho superior aos modelos que se baseiam apenas em informações estruturadas.*

Geralmente, em *Social Commerce*, os produtos comercializados através dos anúncios apresentam imagens, indispensáveis para a conclusão da venda. Afinal, enriquecem a descrição do que está sendo comercializado informando, por exemplo, sobre variações em seu estado (arranhões ou outros problemas) e podem conter indícios da ocorrência de fraude.

1.4 Objetivos do Trabalho

Este trabalho tem como objetivo elaborar um método para detectar fraudes em Social Commerce através das imagens e dados estruturados presentes nos anúncios, através do uso de técnicas capazes de aprender novas representações de forma automática. Especificamente, busca-se a validação da hipótese por meio da utilização de abordagens distintas para a combinação de imagens com dados estruturados dos anúncios. Com isso, espera-se um desempenho superior ao obtido pela utilização de atributos estruturados ou apenas imagens isoladamente.

1.5 Método

No intuito de buscar evidências experimentais quanto à validade da hipótese levantada, o presente trabalho propõe uma metodologia e diversos modelos para medir e avaliar a importância das imagens como fonte de informação para detecção da fraude. Sendo assim, foram construídas arquiteturas e abordagens para análise de atributos estruturados e imagens. Desse modo, são apresentados esquemáticos de cada uma destas, que detalham o procedimento para a combinação.

Para agregar as contribuições das informações contidas nas imagens aos dados estruturados é feito o processamento das imagens. Para este processamento, foram usados algoritmos de aprendizagem profunda¹, do tipo convolucional (CNN), associada à técnica de *Transfer Learning*(28). A rede CNN adotada foi a *MobileNetv2*(29), sua escolha foi devido a sua arquitetura e facilidade em manuseio. Esta rede CNN viabilizou a transformação da imagem em dados estruturados. Com isso, foram propostas duas abordagens, construindo uma base de dados estruturada, em que dados estruturados dos anúncios, mais atributos correspondentes ao resultado dos modelos de classificação das imagens; ou construindo um novo classificador do tipo ensemble², em que o resultado da classificação dos dados estruturados dos anúncios são associados às saídas das classificações das imagens. Somente após uma das duas abordagens que o anúncio foi classificado entre fraude e não-fraude.

Na primeira abordagem, a base de dados estruturada foi enriquecida com os atributos da imagem após o processamento pela rede neural profunda, gerando, especificamente, a probabilidade de cada uma das imagens ser fraudulenta. A agregação destas aos dados estruturados do anúncio incluiu uma informação que foi considerada para a avaliação do algoritmo que classifica os anúncios.

¹algoritmos de Aprendizado de Máquina que inclui extração de características dos dados fornecidos, capazes de aprender ao serem submetidos a dados brutos (27)

²Um *ensemble* é um comitê de classificadores. Comitês de classificadores são compostos por métodos ou algoritmos de classificação e um de combinação de resultados (30) ou de dados, com o objetivo de unir técnicas complementares (ou uma mesma técnica aplicada em bases de dados diferentes) para melhorar o resultado da classificação.

Já na segunda abordagem, são combinadas as saídas do classificador de imagem com as saídas dos classificadores de dados estruturados, que, neste caso, podem ser apenas um classificador ou uma combinação de algoritmos de machine learning, *ensemble* a partir dos melhores resultados obtidos com cada algoritmo. Estes dados combinados, são então avaliados por um algoritmo que irá classificar o anúncio.

Com ambas as abordagens, são exploradas diversas possibilidades de combinação de informações, de modo que, com isso se valide ou refute a hipótese inicial. E, por fim, foram feitas as análises de resultados de cada um dos experimentos, comparando, estatisticamente os resultados obtidos.

Experimentos realizados em uma base de anúncios real de uma grande empresa de *Social Commerce* indicaram ganhos de robustez com os modelos gerados pelo método proposto, obtendo, desta forma, evidências experimentais que apontam para a validade da hipótese levantada. Infelizmente, devido a termo de sigilo assinado, maiores informações sobre a empresa e código desenvolvido não podem ser divulgados.

1.6 Contribuições Esperadas

Entre as principais contribuições do presente trabalho, destacam-se:

- (i) Levantamento bibliográfico amplo sobre o tema de fraude e *e-commerce*
- (ii) Proposta de Modelo Comparativo entre trabalhos relacionados à detecção de fraude em *e-commerce*
- (iii) Metodologia para fusão de dados estruturados e imagens
- (iv) Processo de extração de atributos de imagens para enriquecimento de dados estruturados envolvendo detecção de fraude em *social commerce*
- (v) Protótipo com base no método proposto
- (vi) Evidências experimentais significativas que apontam para a validade da hipótese levantada, na qual existe melhoria no desempenho após a inclusão das imagens nos atributos analisados e, com isso, novas perspectivas para detecção de fraude em ambientes online (não limitados a *social commerce*)

1.7 Organização do Texto

O restante desta dissertação segue organizado da seguinte forma: o Capítulo 2 apresenta, de forma breve, alguns dos conceitos cujo entendimento facilita a compreensão desta pesquisa, já no Capítulo 3 apresenta ampla pesquisa bibliográfica descrevendo

sucintamente os principais trabalhos relacionados; o Capítulo 4 exibe a formalização da metodologia proposta, descrevendo detalhadamente cada uma das etapas envolvidas; o Capítulo 5 apresenta os experimentos realizados e resultados obtidos; e, por fim, o último capítulo aponta as conclusões, destacando as principais contribuições do trabalho, apontando perspectivas de continuidade deste trabalho.

2 CONCEITOS BÁSICOS

A seguir são apresentados conhecimentos para auxiliar as abordagens utilizadas na decorrer desta dissertação.

2.1 Comércio eletrônico (E-commerce)

As TCD (transações comerciais digitais), são transações comerciais que ocorrem em um meio digital (i.e.: internet) assistida por uma TIC (Tecnologia de Informação e Comunicação). Sendo assim, é importante detalhar o *e-commerce* como sendo a troca de bens e serviços entre organizações independentes e/ou pessoas, suportadas por sistemas TIC numa infraestrutura de rede (31, 32, 33).

Contudo, o *e-commerce* pode ocorrer em diversos locais que realizam TCD, como lojas *online* e *marketplace*, além destes destaca-se também os classificados *online* e Leilão *online*.

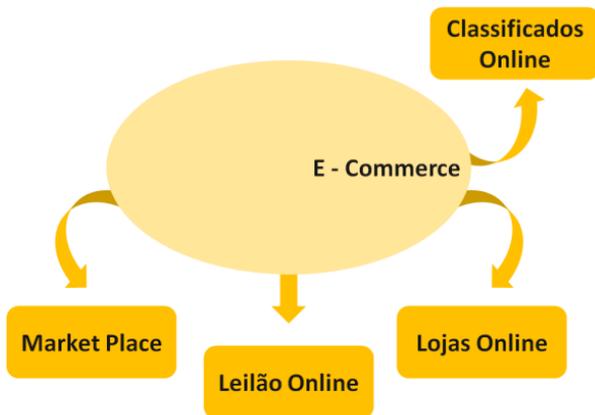


Figura 3 – Representação de locais onde ocorrem e-commerce

Até onde foi possível identificar na bibliografia, não foi encontrado uma proposta quanto à organização estrutural/relacional destes. Salvo (31) que aponta *marketplace* como um local *online* onde fornecedores e consumidores em potencial se encontram para fazer TCD. Desse modo, agregando à identificação da lacuna os conhecimentos obtidos das pesquisas aos conhecimentos de domínio sobre o assunto, é proposta, visando facilitar o entendimento do leitor, a seguinte taxonomia:

- *E-Commerce*

1. Lojas *online*
2. *Marketplace*
 - a) Classificados *online*
 - b) Leilão *online*

Assim, nos classificados *online* e leilões *online*, vendedores e compradores alternariam suas posições entre fornecedores e compradores conforme seus interesses e necessidades, ainda que tenham a opção de permanecer perenemente em apenas uma posição.

Desse modo, é apontada, na proposta, que a principal diferença entre Lojas *online* e *marketplace* é a relação comprador-fornecedor. Onde as Lojas *online* seriam caracterizadas por relações (1-1) enquanto que no *marketplace* são observadas relações (1-n ou n-n). E, portanto, classificados *online* e leilões *online* seriam, por esta proposta, um *marketplace*; assim como, duas ou mais lojas *online* também comporiam um *marketplace*. Ressaltando-se a necessidade de que neste mesmo local existam duas ou mais lojas, pois, do contrário, essa relação seria (1-1), devendo, portanto, ser classificada como uma loja *online*.

Além destes, há o conceito de *Social Commerce* (*S-commerce*), que é considerado um subconjunto do *e-commerce* (13, 34), composto de dois elementos chave:

1. Redes Sociais
2. Atividades comerciais

Conforme é possível observar na Figura 4 e o maior aprofundamento do tema, é percebido que, em sua maioria(13), são a incorporação de:

- i. Características de atividades comerciais às Redes Sociais, deixando de apenas conectar pessoas para realizar TCD
- ii. Ferramentas e funcionalidades de rede social ao *e-commerce*, visando facilitar a interação social e o compartilhamento de informações

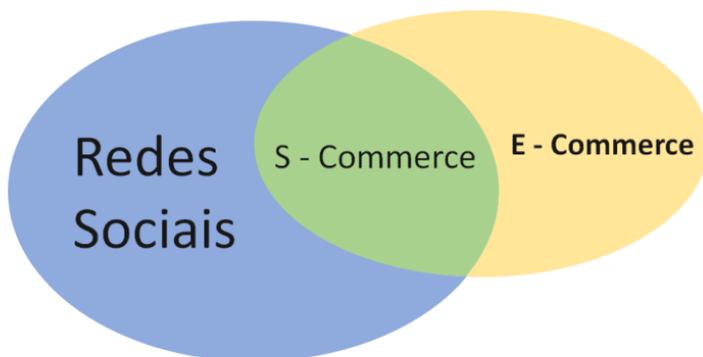


Figura 4 – Redes Sociais + E-commerce = Representação Visual do Social Commerce

A inclusão do *Social Commerce* à taxonomia proposta analisa suas relações (1-n ou n-n), e portanto, classifica junto ao *marketplace*. Entretanto, nem todo *Social Commerce* é um *marketplace*, porém todo *marketplace* é um *Social Commerce*, dessa forma, a nomenclatura de *marketplace* pertence ao *Social Commerce*.

- *E-Commerce*
 1. Lojas *online*
 2. *Social Commerce*
 - a) *Marketplace*
 - i. Classificados *online*
 - ii. Leilão *online*

2.2 Definições de Fraude

Complementando as definições de fraude e impactos ao indivíduo da introdução, é exposta parte da pesquisa feita por (3). Dessa forma, tendo como pré-requisito a perda financeira sofrida por uma vítima, são explicados os tipos de fraude.

Os alvos a serem fraudados podem ser indivíduos ou organizações (grupos de indivíduos) conforme Figura 5.

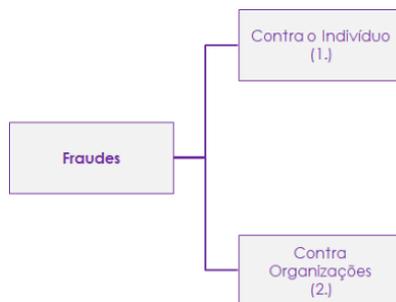


Figura 5 – Detalhamento de Alvos de Fraude (3)

Fraudes cometida contra o indivíduo, podem ser classificadas desde 1.1 até 1.7. Em cada uma delas, o importante é que o fraudado ele sempre espera receber algum retorno, conforme Figura 6.

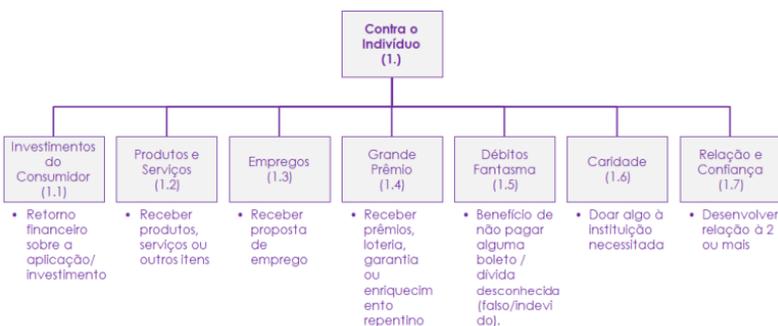


Figura 6 – Detalhamento de Fraude Contra o Indivíduo (3)

O problema que será estudado encontra-se na Fraude contra o Indivíduo (1.), Produtos e Serviços (1.2.).

A taxonomia também define o que seriam fraudes contra organizações, que podem ser contra a sociedade, agências governamentais reguladoras, ou organizações e negócios de quaisquer natureza, conforme Figura 7.

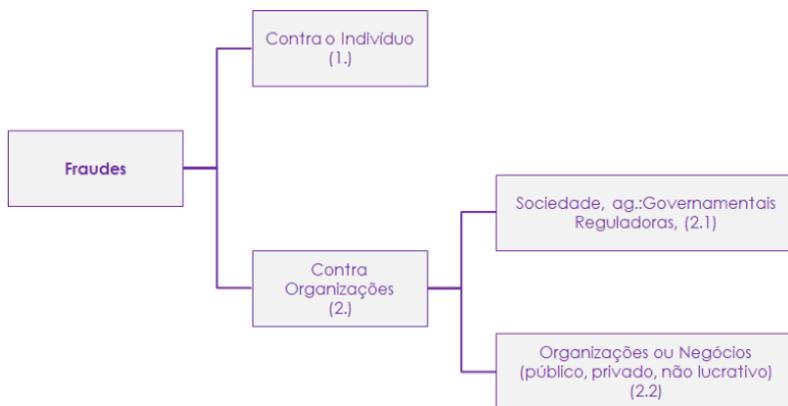


Figura 7 – Detalhamento de Fraude Contra Organizações (3)

Apesar da taxonomia ter maior profundidade, com o exposto é possível garantir alinhamento fraude para finalidade desta dissertação, bem como quem é o alvo.

2.3 Aprendizado de Máquina

As abordagens utilizadas para o problema de detecção de fraude se dividem em supervisionada e não supervisionada. Na abordagem não supervisionada, não é necessário o conhecimento prévio da saída desejada, diferentemente do aprendizado supervisionado, onde os dados são conhecidos e é possível auxiliar o algoritmo a obter o melhor resultado possível, conforme observado na Figura 8.

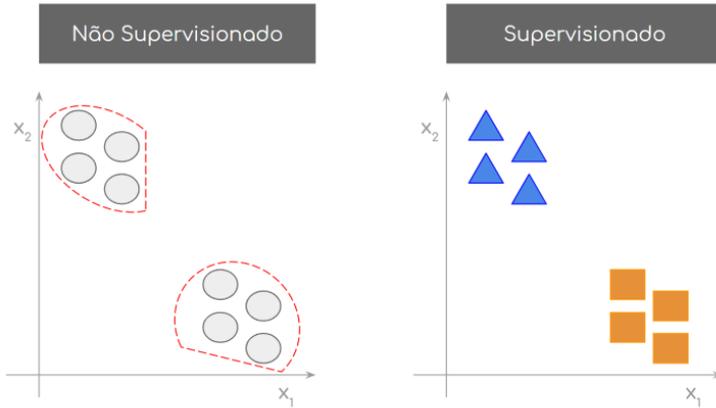


Figura 8 – Ilustração aprendizado supervisionado e não supervisionado

2.3.1 Supervisionado

No aprendizado supervisionado, destaca-se a utilização das técnicas de Classificação, Regressão e Previsão de Séries temporais, onde, por meio de dados previamente rotulados, deseja-se estimar o rótulo de saída.

Na Classificação se trabalha com rótulos categóricos, como por exemplo, as chances de um usuário ser fraude. Baseado nos dados deste usuário, é estimada uma chance deste usuário ser fraude, que no final, por meio de uma função de ativação, será classificado como fraude ou como idôneo.

Na Regressão, onde são explorados rótulos contínuos, similar à classificação, analisando dados deste usuário, estima-se qual poderia, por exemplo, ser a nota que um indivíduo irá tirar numa prova, baseado em seu índice de absentéismo e número de perguntas feitas em sala de aula.

Ainda na linha de aprendizado supervisionado, existe uma aplicação que envolve a previsão de séries temporais, onde o rótulo, assim como na regressão, é contínuo. Com a particularidade de ser dependente do tempo. Somado a isso, outra diferença entre a Regressão e a Previsão de Séries temporais é a utilização dos dados de saída para prever novos dados de saída. Ou seja, diferente da nota do aluno, que é estimada por parâmetros alheios à este, são utilizadas as próprias notas históricas deste, para então prever qual poderiam ser as próximas notas deste aluno.

2.3.2 Não Supervisionado

O aprendizado não supervisionado, por sua vez, é caracterizado, principalmente pelo desconhecimento dos rótulos de saída. Por isso, aplicar as diversas técnicas de aprendizado de máquina torna possível lidar com dados pouco conhecidos e observar características em comum entre eles.

Explorando esta modalidade de aprendizado, é possível destacar a técnica de agrupamento, que permite, por meio da aproximação de pontos a identificação de semelhanças e grupos entre registros. Iniciando por meio da observação de cada um dos pontos e então traçando grupos de tamanhos previamente estipulados, desde que eles estejam a uma distância entre si também previamente estipulada, por exemplo.

Além disso, há também técnicas de associação, onde são descobertas relações entre variáveis. Esse tipo de técnica se tornou bem popular em segmentos do setor *online* de venda de produtos. Onde, baseado na escolha feita por um indivíduo, são ofertados novos produtos que ele teria uma maior propensão a levar (baseado em histórico de aquisição de outros clientes de perfis similares).

2.3.3 Técnicas de Mineração de Dados

Uma outra vertente que surge, embora seja muito citada em conjunto às técnicas de aprendizado de máquina, são as técnicas de mineração de dados. Cuja principal semelhança é a proposta em responder questionamentos baseado em dados e não em sensações ou percepções que podem conter viés.

Para ilustrar um pouco sobre às formas de lidar com problemas de modo automatizado, percebe-se que as tarefas de mineração de dados, seguem a mesma estrutura lógica do aprendizado de máquina, podendo ser agrupadas em:

- Regressão
 - Predição do Valor das Vendas em uma nova Filial
- Previsão
 - Previsão de Vendas em Varejo
- Associação
 - Oferta de Novos Serviços e Produtos (Netflix, Amazon, Americanas.com, etc.)
- Agrupamento
 - Determinação de Localidades Promissoras para Abertura de Novas Filiais

- Classificação
 - Detecção de Clientes com Perfis Fraudulentos

Neste trabalho será utilizada a classificação, pois é desejado descobrir se a nova transação é uma fraude ou não.

2.3.3.1 Classificação - Técnicas tradicionais

Existem muitas técnicas para lidar com a classificação, é possível observar algumas delas aqui (35):

- Baseadas em Instância - K-NN
- Árvores de Decisão - ID3, C4.5
- Estatística - Classificadores Bayesianos
- Redes Neurais - MLP ou Aprendizado Profundo

2.3.4 Aprendizado Profundo

Devido a sua complexidade e por ser muito adotada ao longo deste trabalho, foi criado um tópico separado para este item.

Redes de aprendizado profundo, em inglês (deep learning) têm sido adotadas para extração de atributos devido a sua capacidade de organizar e entender o mundo por meio de uma hierarquia de conceitos(36), desse modo permitindo que conceitos complexos sejam entendidos por meio da construção de diversas camadas de conceitos mais simples. Tais métodos tem melhorado consideravelmente o estado da arte em reconhecimento de fala, reconhecimento visual e detecção de objetos, por exemplo(27).

Uma das formas de utilizar redes profundas é por meio da técnica de *Transfer Learning*, que pode ser definido como um aprendizado melhorado, uma vez que se beneficia da transferência de um conhecimento relacionado previamente adquirido(28). Não somente por estes benefícios como sua utilização é recomendada quando se tem uma tarefa de classificação em determinado domínio de interesse (i.e. fraude); porém, só há dados disponíveis em outros domínios (37).

3 TRABALHOS RELACIONADOS

Uma extensa busca foi feita usando os termos do tema proposto 'detecção de fraudes em *social commerce*'. Contudo, foram retornados poucos artigos. Conforme a pesquisa era aprofundada, por meio da leitura dos artigos recuperados pelas ferramentas de busca, somados aos artigos referenciados por estes, entendeu-se a necessidade de expandir a busca. Para esta nova pesquisa houve a necessidade de incluir outros termos similares à *social commerce*, conforme fora detalhado no capítulo 2, além de expandir também a pesquisa às diversas modalidades de fraude existentes e os cenários que elas costumam acontecer. Desse modo, neste capítulo, foi feita uma organização dos trabalhos encontrados desde 1995, consolidados pela Tabela 1.

Tabela 1 – Quantidade de Trabalhos sobre Fraude desde 1995

Ano	Crédito	Automobilística	Telecomunicações	Leilão Online	Plano Saúde	Total
1995		01				01
1997	02	01	01			04
1998		02	01			03
1999	01		01			02
2000	01			01	01	03
2001		01	01			02
2002		04	01	01		06
2003	01	01	01	01	01	05
2004	01	01	01	02	01	06
2005	01	06		01		08
2006	01	01	03	01	03	09
2007		01	02			03
2008	03	02			01	06
2009	02	01	04	02		09
2010	02			02	01	05
2011	05	03	02	04	02	16
2012	04	02	02	03		11
2013	04		01	01	02	08
2014	02		01	02	01	06
2018	02					02
2019				01		01
Total	32	27	23	23	13	118

Ao longo da construção da Tabela 1, percebeu-se que, neste tema, os dados costumam ser sensíveis e raramente divulgados. O que pode ser entendido como uma questão estratégica das empresas - expor esta informação pode prejudicar o negócio ao ser explorada por pessoas mal intencionadas.

Contudo, observou-se que a pesquisa para a criação de modelos de detecção de fraudes em transações de vendas de *e-commerce*, vem se desenvolvendo há vários anos (38, 39, 40, 41, 42) e abrange diversos contextos de aplicação tais como, por exemplo, leilões *online* (43, 44), *marketplace* (45, 46), redes sociais (47, 48, 49, 50, 51) e, mais recentemente, *social commerce*. Embora existam aplicações que possam ser classificadas em mais de um cenário, o presente trabalho se concentrou prioritariamente em *social commerce*, devido à

sua crescente expansão potencializada pelas redes sociais.

Até onde foi possível observar, diante da revisão da literatura especializada, destaca-se que os principais trabalhos voltados à detecção de fraude, em anúncios de vendas de produtos oferecidos por meio de *Social Commerce*, utilizam modelos de classificação binária (fraude x não fraude) construídos pela aplicação de algoritmos de aprendizado de máquina.

3.1 Modelo de Dados Conceitual de *Social Commerce*

Para maior clareza tanto sobre o *social commerce* quanto sobre a forma com que os relacionamentos entre seus elementos são estabelecidos foi necessária uma modelagem. A construção desta em paralelo à leitura e estudo dos trabalhos selecionados permitiu uma maior clareza sobre oportunidades de exploração de informações existentes.

Por meio da análise da proposta de modelo de dados conceitual, apresentado na Figura 9, foi possível compreender que a entidade anúncio é a peça chave que permite a conexão entre diversos usuários. Como parte das características de um anúncio, observam-se atributos estruturados (data de publicação, categoria) e atributos não-estruturados (conteúdo textual do título, texto descritivo do produto anunciado, imagens do produto anunciado). O anúncio também viabiliza a troca de mensagens entre usuários, referente a um anúncio específico, a conclusão da negociação com a efetiva compra. Outra entidade que foi percebida como de grande importância foi a entidade Indivíduo (pessoa), seja ela a pessoa que esta publicando um anúncio ou a pessoa buscando por produto específico.

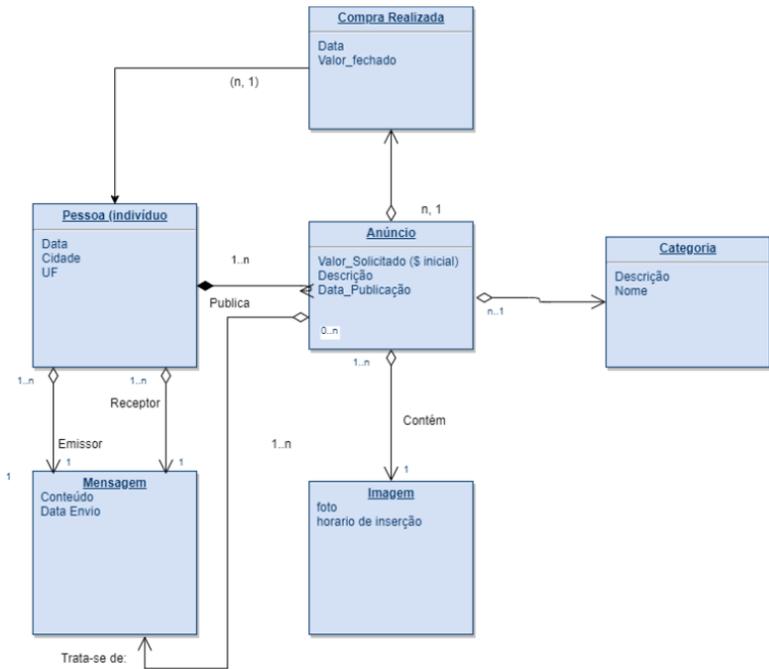


Figura 9 – Proposta de Modelo de Dados Conceitual

3.2 Proposta de Modelo Comparativo

Com o objetivo de estruturar os conceitos e permitir ao leitor um maior entendimento do assunto, é proposta uma distribuição dos trabalhos relacionados conforme a revisão bibliográfica. Esta organização foi feita a partir do ano de 2012, onde fraude ganha uma repercussão maior.

3.2.1 Organização conforme tópicos

Diferente do modelo de dados conceitual, a proposta da apresentação dos trabalhos relacionados não tinha como objetivo ser específica e restrita ao universo de *social commerce*. Dessa forma, os trabalhos foram organizados de modo a permitir a comparabilidade entre trabalhos relacionados desenvolvidos para aplicações distintas (Leilões online, Crédito, Telecomunicações, Planos de Saúde, Automobilística).

A partir dessa organização foi possível a análise dos atributos mencionados nos trabalhos em função do tipo de informação (estruturada ou não estruturada) e da origem da informação (separados entre entidade e características). Com esta organização foi possível detalhar cada um dos atributos independente do ambiente ao qual ele se aplica e comparar os trabalhos. Também com esta organização e maior detalhamento dos trabalhos, foi possível selecionar os trabalhos com maior afinidade com o que será explorado nesta dissertação. Assim, os atributos foram segmentados conforme o tipo de informação e origem da informação.

3.2.1.1 Classificação em função do tipo de informação explorada

Nesta segmentação os atributos explorados nos artigos são organizados de acordo com a organização e natureza dos dados, que pode ser:

- Estruturada
- Não-Estruturada
 - Texto
 - Imagem
 - Áudio

Por exemplo, pode-se ter dados estruturados tais como: número de dias desde a última transação ou valor da transação. Mas por outro lado, pode-se ter como dado não estruturado: a gravação de uma ligação reportando um incidente, uma imagem que contém informações sobre o incidente (seguro), sobre o produto que está sendo leiloado, ou ainda textos que descrevam todo o relato do ocorrido ou o laudo médico descrevendo todos os procedimentos executados por um médico. Todas essas informações seriam dados Não-estruturados, envolvendo texto, imagem ou áudio.

3.2.1.2 Classificação em função da origem da informação

A organização conforme a origem dos elementos descritos, permitiu o agrupamento de atributos em Objetos, Indivíduos, Ações, Comunidade. Para tal, foram consideradas as características atuais [CA], que descrevem o estado atual dos elementos, ou características históricas [CH], responsáveis pelo armazenamento de estados passados, registrando suas mudanças e alterações.

Com isso, foi obtido o mapeamento a seguir. Nas próximas subseções, foram descritos e exemplificados cada um dos elementos (ou entidades) e suas características:

- Objetos

- Características Atuais [CA]
- Características Históricas [CH]
- Indivíduos
 - Características Atuais [CA]
 - Características Históricas [CH]
- Ações
 - Ações Usuário-Usuário (UU)
 - * Características Atuais [CA]
 - * Características Históricas [CH]
 - Ações Usuário-Produto (UP)
 - * Características Atuais [CA]
 - * Características Históricas [CH]
- Comunidade
 - Comparação do Objeto com o todo
 - Comparação do Indivíduo com o todo
 - Comparação das Ações Usuário-Usuário com o todo
 - Comparação das Ações Usuário-Produto com o todo

3.2.1.2.1 Objetos

O agrupamento conforme a entidade objeto considera atributos que qualifiquem o objeto, produto ou serviço, que está sendo transacionado entre as partes. Esta entidade, em *social commerce* representada pelo anúncio, permite que ambos os envolvidos no processo, vendedor e comprador, se conectem. Na tabela 2, podem ser observados alguns exemplos.

Tabela 2 – Exemplo de Agrupamento das Características dos Objetos

Ambiente	Contexto	Objeto	Exemplos de atributos	Estruturado?
Leilão online	Produto / Serviço em negociação	Anúncio de Playstation4	Tempo de Uso[CA], Cor[CA], Imagem do Produto[CA], Descrição Atual[CA], Descrição Prévia[CH]	Sim, Sim, Não, Não, Não
Crédito	Aquisição de item online	Ingresso	Local do evento[CA], Horário[CA], Gênero da Atração[CA], Descrição do evento (Campo Textual)[CA]	Sim, Sim, Sim, Não
Telecomunicações	Ligação para Cancelamento de serviços contratados	Ligação telefônica	Duração[CA], Nacional/internacional[CA], Áudio da Ligação[CA], Horário de início[CA]	Sim, Sim, Não, Sim
Planos de Saúde	Serviço de raio-X em solicitação de reembolso	Serviço Realizado	Valor do serviço[CA], Anotações Médicas[CA]	Sim, Não
Automobilística	Acionamento da franquia	Carro assegurado	Ano do carro[CA], quantidade de portas[CA], valor na tabela FIPE[CA], Imagem de Cadastro[CH], Imagem do Sinistro[CA]	Sim, Sim, Sim, Não, Não

3.2.1.2.2 Indivíduos

O agrupamento conforme a entidade Indivíduos, considera o ambiente estudado e o realizador da ação. Dessa forma, estes podem ser representados, na prática, por pessoas, cartões de crédito, etc. Na tabela 3, podem ser observados alguns exemplos.

Tabela 3 – Exemplo Agrupamento dos atributos em Indivíduos

Ambiente	Contexto	Indivíduo	Exemplos de atributos	Estruturado?
Leilão online	Produto / Serviço em negociação	Comprador ou Vendedor	Tempo de Conta[CA], Quantidade Vendida[CH], Foto Perfil[CA]	Sim, Sim, Não
Crédito	Aquisição de item online	Cartão de Crédito	Moeda[CA], Paga em dia?[CH], Foto da Assinatura[CA]	Sim, Sim, Não
Telecomunicações	Ligação para Cancelamento de serviços contratados	Número da Linha Telefônica	Quantidade Ligações[CH], Duração média[CH], Turno Frequente[CH]	Sim, Sim, Sim
Planos de Saúde	Serviço de raio-X em solicitação de reembolso	Paciente, Médico, Hospital	valor médio[CH], recência[CA], frequência[CH], Foto Crachá[CA]	Sim, Sim, Sim, Não
Automobilística	Acionamento da franquia	Motorista Assegurado ou Envolvidos (acidente)	Gênero[CA], Idade[CA], Renovações CNH[CH], Foto CNH[CA]	Sim, Sim, Sim, Não

Destaca-se aqui que numa venda à crédito, o Indivíduo é o cartão e não o proprietário do cartão. Tal modelagem é feita porque entende-se que um mesmo proprietário pode ter comportamentos diferentes com cada cartão, tendo um para cada finalidade.

Também ressalta-se que existem modelagens com múltiplos indivíduos num mesmo ambiente/contexto. Como, por exemplo, num leilão online composto por dois ou mais indivíduos, a conta do anunciante na plataforma e a conta do comprador. Por outro lado, nos planos de saúde, cada entidade envolvida pode ser modelada como um indivíduo, uma vez que todos estes podem configurar fraude (i.e.: Médico que passa a conta em um hospital que nunca atendeu, ou que não existe, ou quando um médico específico faz a requisição de um processo superfaturado). Já para o setor da automobilística são considerados como indivíduos tanto o assegurado quanto os envolvidos.

3.2.1.2.3 Ações - Usuário/Usuário(UU) ou Usuário-Plataforma(UP)

As entidades definidas como Ações estão relacionadas às atividades que são o agrupamento em função das ações realizadas no ambiente estudado. Todas as ações se limitam à ações entre indivíduos (UU) ou indivíduo e plataforma (UP). Na tabela 4, podem ser observados alguns exemplos.

Tabela 4 – Exemplo Agrupamento dos atributos em Ações

Ambiente	Contexto	Ações (UU / UP)	Exemplos de atributos	Estruturado?
Leilão online	Produto / Serviço em negociação	Realizar lance(UP), enviar mensagem(UU)	Valor do Lance[CA], Tamanho Mensagem[CA], Texto Mensagem[CA]	Sim, Sim, Não
Crédito	Aquisição de item online	Digitar CVV(UP), Mudança Pais(UP), Captcha(UP)	erro de CVV[CA], conversão de moeda?[CA], Imagem Captcha[CA]	Sim, Sim, Não
Telecomunicações	Ligação para Cancelamento de serviços contratados	Digitar Telefone(UP), Gravar Áudio (UP)	Tempo de Digitação[CA], Áudio da gravação[CA]	Sim, Não
Planos de Saúde	Serviço de raio-X em solicitação de reembolso	Acessar Aplicativo (UP)	cliques em Doutores[CH], Foto Doutor[CA]	Sim, Não
Automobilística	Acionamento da franquia	Ligar solicitando Reboque (UP)	Endereço[CA], Foto CNH[CA]	Sim, Não

3.2.1.2.4 Comunidade

A Comunidade é caracterizada por conter informações dos Objetos, Indivíduos ou Ações avaliados coletivamente. Desse modo é possível gerar métricas comparativas (ex: média, desvio padrão) em relação à população total ou parcial analisada. Na tabela 5, podem ser observados alguns exemplos.

Tabela 5 – Exemplo Agrupamento dos atributos em Comunidade

Ambiente	Contexto	Comunidade	Exemplos de atributos	Estruturado?
Leilão online	Produto / Serviço em negociação	Objeto (Playstation)	Tempo de uso comparado com outros Playstation	Sim
Crédito	Aquisição de item online	Ação (Inserção de Código CVV)	Média diária de erros CVV	Sim
Telecomunicações	Ligação para Cancelamento de serviços contratados	Indivíduo (Número da Linha Telefônica)	Desvio padrão da Duração Média	Sim
Planos de Saúde	Serviço de raio-X em solicitação de reembolso	Objeto (Serviço Realizado)	Mediana do serviço	Sim
Automobilística	Acionamento da franquia	Indivíduo (Motorista em acidente)	Histograma de acidentes agrupados por idade	Sim

Do ponto de vista da avaliação coletiva, pode-se analisar a quantidade total ou parcial de algum tipo de entidade. Por exemplo, num leilão online, para uma análise de preço de objeto seriam obtidos preço comparado com todos os outros preços existentes no site, mas também poderiam ser comparados apenas com preços da categoria, do dia, produtos de mesmo nome, etc.

3.3 Aplicação do Modelo Comparativo

Para organizar os trabalhos conforme o modelo comparativo, é necessário que seja possível analisar claramente os atributos utilizados por estes autores; porém, em alguns dos trabalhos levantados os autores não deixaram minimamente explícito os atributos utilizados. Por este motivo, a Tabela 6 apresenta os trabalhos descartados por este motivo.

Tabela 6 – Autores com informações não explícitas

Ano	Autor	Não explícitas
2012	Akhter e Ahamad(52)	1
2012	Brennan e Hoffman(53)	1
2012	Phua et al.(54)	1
2013	Dal Pozzolo et al.(55)	1
2018	Carcillo et al.(56)	1

Os trabalhos que puderam ser detalhados são apresentados na Tabela 7.

Tabela 7 – Publicações de Fraude por Ano desde 2012

Ano	Autor	Objeto [CA],[CH]	Indivíduos [CH]	Indivíduos [CA]	Ações (UU,UP)[CA]	Ações (UU,UP)[CH]	Comunidade Objetos	Comunidade Indivíduos	Comunidade Ações
2012	Augustin et al.(57)	1	1	1					
2012	Chang e Chang(58)		1			1	1		
2012	Ford, Xu e Valova(23)	1		1	1				1
2012	Jha, Guillen e Christopher Westland(59)	1	1	1		1		1	
2012	Lei(60)	1		1	1				
2012	Han Tao, Liu Zhixin e Song Xiaodong(61)	1		1	1				
2012	Tsang, Dobbie e Koh(62)		1	1					1
2012	Wong et al.(63)			1	1				
2013	Al-Fayoumi e Shilbayeh(64)		1	1					
2013	Almendra(22)	1	1	1			1	1	
2013	Chen e Gangopadhyay(65)							1	1
2013	Liu et al.(66)	1	1	1					
2013	Malekian e Hashemi(67)		1	1					
2013	Sahin, Bulkan e Duman(68)		1	1				1	
2014	Chang e Chang(69)		1		1	1			1
2014	Dal Pozzolo et al.(70)		1					1	
2014	Saravanan et al.(71)		1	1				1	
2014	Seeja e Zareapoor(72)	1		1					
2014	Tsai, Ko e Lin(73)	1							
2014	Tsang et al.(74)		1		1	1		1	1
2018	Cheung, She e Liu(20)	1							
2018	Zanin et al.(75)	1	1	1				1	
2019	Carta et al.(21)	1							

A análise da tabela mostrou uma grande concentração dos trabalhos nos Indivíduos, tanto em características históricas quanto atuais. Desse modo, foram percebidas oportunidades para aprofundar a abordagem na ação e no objeto. Dada essa lacuna, nesta dissertação, optou-se por explorar aprofundar nos trabalhos que exploravam as características do objeto.

Com base no universo mais restrito, considerando apenas trabalhos que analisassem os Objetos, avaliou-se os trabalhos por autores com maior afinidade ao tema proposto, separando os trabalhos em dois grupos em função do tipo de informação utilizada, conforme apresentado na Tabela 8.

No primeiro grupo estão os trabalhos que se baseiam exclusivamente em informações estruturadas extraídas dos anúncios, tais como preço, tipo de produto, marca, se é novo

ou usado, local físico de venda, entre outras (21, 22, 23). O segundo grupo, por outro lado, contém os trabalhos que, analisam o conteúdo das imagens sobre os produtos oferecidos nos anúncios, em busca de indícios de fraude (24, 25, 20, 26). Os próximos parágrafos apresentam sucintamente cada um dos trabalhos indicados.

Tabela 8 – Visão Resumida dos Trabalhos Relacionados

Referência	Dados Estruturados	Imagens	Algoritmos e Técnicas
(21)	sim	não	<i>MLP, GNB, RF, ADA, GBC</i>
(22)	sim	não	<i>RF, LGR, DCT, MLP, SVM, KNN, GBC, Boosted Trees</i>
(23)	sim	não	Rede Neural (adaptativa e com 3 camadas)
(24)	não	sim	atributos textuais <i>OCR</i> obtido das imagens <i>edge based clustering</i> histogramas, <i>layout masks</i>
(25)	não	sim	atributos textuais <i>OCR</i> obtido das imagens <i>edge based clustering</i> histogramas, <i>layout masks</i>
(26)	não	sim	<i>CNN, K-means</i> Gráficos Sociais,
(20)	não	sim	<i>CNN, K-means</i>

Em (21), os autores exploram a utilização de dados estruturados envolvidos nas transações. Para a detecção de transações fraudulentas foram criados *ensembles*, combinando diferentes algoritmos de classificação. Cabe destacar que o texto não deixa explícito se foi realizada algum tipo de seleção de atributos antes da aplicação dos *ensembles*.

Aplicado em dados dos anúncios de uma grande empresa que atua em *social commerce*, o trabalho descrito em (22) utiliza atributos de anúncios (e.g., preço, data de publicação, categoria do anúncio) e do usuário (e.g., tempo de conta, reputação da conta, número de transações recentes) na construção dos modelos de detecção de fraudes. Entre as técnicas de pré-processamento aplicadas, os autores citam técnicas de *downsample* para minimizar o efeito do desbalanceamento entre as classes fraude e não fraude.

Já em (23), os autores procuraram utilizar informações sobre as ações executadas pelos usuários na plataforma computacional durante as transações, de forma a agrupá-los por similaridade de comportamento nos leilões de oferta de produtos. São exemplos de informações utilizadas: tempo desde o anúncio até o lance do comprador, resposta do vendedor e tempo restante até o último lance. O trabalho também explorou o uso de informações sobre o leilão de cada produto (e.g., média de incremento dos lances, número

de lances, etc.) e atributos do vendedor, tais como lance inicial e taxa de resposta. Uma vez agrupados em função da similaridade de comportamento em leilões anteriores, o trabalho buscou identificar usuários suspeitos de fraude a partir de grupos com alta incidência de usuários já historicamente identificados como fraudadores.

Diferentemente dos trabalhos anteriores, em (25), os autores exploram as características presentes nas imagens dos produtos anunciados. Para tanto, utilizaram o conceito de preferências de edição de imagem como um atributo para identificação de autoria e perfil do vendedor em ambientes *online*. Por ser um conceito inovador para a época, introduzido por este trabalho, o autor se concentrou em detalhar o conceito, que consiste em estender a análise da imagem, originalmente concentrada nas cores presentes e na análise das formas, considerando suas bordas e limites. A título de exemplo, caso a imagem incluía caracteres textuais, a análise incluiria informações sobre a posição deste texto na imagem, a fonte usada e a proporção do texto em relação ao objeto apresentado na figura. Por meio de uma sequência de experimentos, os autores foram capazes de desenhar um solução, na qual foi possível a identificação de usuários apenas considerando essa análise estendida das imagens que, comparada com outras técnicas, provou ser melhor para identificação dos vendedores, e, por esse motivo, tornou-se uma referência a ser adotada em trabalhos futuros para identificação de autoria.

(24) aprofunda os estudos analisando os mesmas imagens utilizadas nos experimentos realizados em (25). Neste trabalho, o objetivo principal foi a identificação preventiva de invasão de conta e fraudes na plataforma computacional. A solução apresentada pelos autores se mostrou interessante porque se propunha a avaliar a intenção do usuário, que inseria conteúdo fraudulento na plataforma, e identificar possíveis vítimas de invasão de contas, conduzindo a solução para um nível mais elevado de abstração, além da simples identificação de conteúdo fraudulento nas transações. Para tanto, o trabalho processou cada imagem de acordo com as preferências de edição de imagem do vendedor. Com essa abordagem, os autores conseguiram identificar usuários que cometiam fraude na plataforma, usuários que burlavam as regras da plataforma (tinham múltiplas contas) e as publicações que foram feitas em contas idôneas, vítimas de invasão de conta.

Nos trabalhos descritos em (26) e em (20), os autores também buscaram soluções envolvendo a análise de imagens. No primeiro, a ideia foi identificar conexão de usuários de redes sociais por meio de imagens. Por exemplo, uma mesma foto de um grupo de pessoas, relacionadas em uma rede social e compartilhada no perfil desses usuários, pode ser um forte indicativo de que eles têm conexões entre si. Os resultados obtidos nos experimentos realizados confirmaram a adequação da estratégia utilizada pelos autores para identificar associações entre usuários de redes sociais. Embora não relacionado diretamente à detecção de fraude, este primeiro trabalho deu origem a um segundo estudo, do mesmo autor principal, no qual foi investigada a descoberta de associações entre imagens para descobrir

fraudadores (mais especificamente falsificadores atuando como potenciais vendedores) nas redes sociais *Instagram* e *Carousell*. Em linhas gerais, a ideia do trabalho foi identificar grupos de usuários com perfil de imagens compartilhadas semelhantes entre si. A partir desses grupos, de forma análoga à estratégia adotada em (23), os autores identificaram vendedores suspeitos a partir de grupos com elevada incidência de fraudadores conhecidos.

Diante do exposto, pode-se perceber, até onde foi possível na pesquisa bibliográfica realizada, que nenhum dos trabalhos relacionados investigou a combinação de informações estruturadas dos anúncios com informações extraídas das imagens associadas a esses anúncios. Desta forma, todos eles deixaram de explorar fontes de informação (dados estruturados ou imagens) com possíveis indícios de fraude.

4 MÉTODO PROPOSTO

Denominado Detecção de Fraude em *Social Commerce* (DFSC), o método proposto neste trabalho permite combinar informações estruturadas com informações extraídas de imagens de anúncios de produtos comercializados em ambientes de *Social Commerce*, a fim de detectar a presença de fraude nesses anúncios. O DFSC foi desenvolvido de forma a oferecer quatro alternativas de utilização independentes: DFSC(Est), DFSC(Img), DFSC(Est+Img) e DFSC(Est,Img), todas ilustradas na Figura 10.

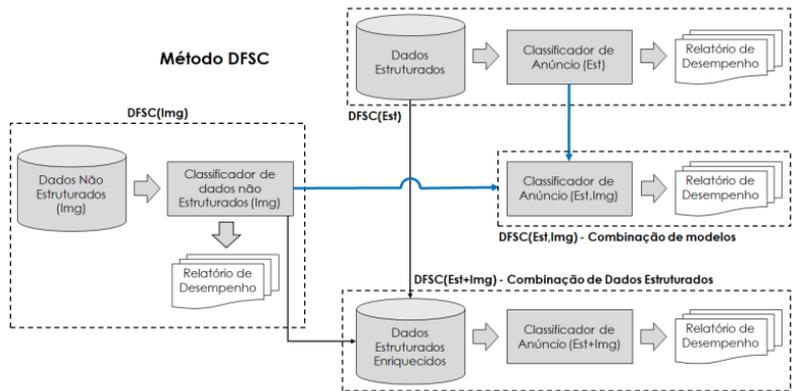


Figura 10 – DFSC: Visão Geral das Alternativas de Aplicação

Para a alternativa DFSC(Est), o método proposto considera apenas informações estruturadas sobre os anúncios para detectar fraudes. Tal alternativa tem como objetivo permitir a construção e execução de modelos de classificação, similares aos existentes no estado da arte (i.e., modelos de detecção de fraudes que se baseiam exclusivamente em informações estruturadas existentes nos anúncios).

Na alternativa DFSC(Img), o método proposto constrói e aplica modelos que consideram exclusivamente a imagem de um anúncio a fim de detectar indícios de ocorrência de fraude em tal imagem. As demais alternativas são abordagens que, de fato, combinam informações estruturadas com informações de imagens a fim de detectar fraude.

De forma resumida, a abordagem de combinação DFSC(Est+Img) constrói e aplica primeiro um modelo de classificação de imagem para extrair informações das imagens e enriquecer os dados estruturados do anúncio com essas informações. Em seguida, constrói

e aplica um modelo de classificação sobre os dados enriquecidos a fim de classificar o anúncio como possível fraude.

Por outro lado, a abordagem de combinação DFSC(Est,Img) constrói e aplica inicialmente dois modelos de classificação independentes, sendo um sobre os dados estruturados do anúncio e outro sobre os dados das imagens desse anúncio. Em seguida, os resultados gerados por esses modelos são, então, submetidos a um terceiro modelo classificação que tem como objetivo apontar se o anúncio em questão pode ou não conter indícios de fraude. Os detalhes das especificidades de funcionamento e da construção dos modelos de classificação das quatro alternativas encontram-se nas Seções 4.1, 4.2, 4.3 e 4.4. A Figura 11 e os próximos parágrafos descrevem as etapas comuns às alternativas de utilização do DFSC.

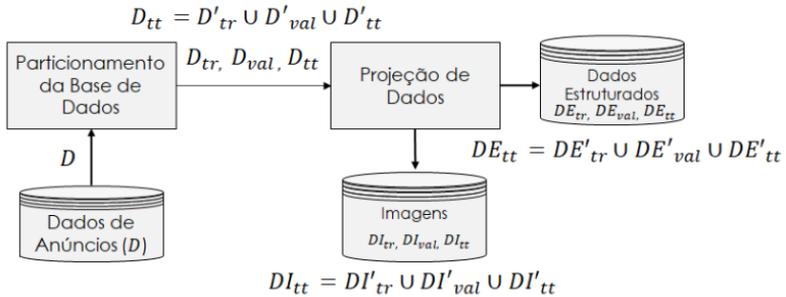


Figura 11 – Visão Geral das Etapas Comuns à Utilização do DFSC

Inicialmente, o DFSC recebe como entrada uma base de dados D composta por anúncios, onde cada anúncio $a = (e_{1,a}, e_{2,a}, \dots, e_{m,a}, c_a, I_a) \in D$ possui informações estruturadas $e_{1,a}, e_{2,a}, \dots, e_{m,a}$ (representadas, neste trabalho, de forma simplificada por e_a) tais como, por exemplo, preço, marca do produto, entre outros. Além disso, a possui c_a , um rótulo que indica se a é fraude ou não (duas classes de saída), e $I_a = \{(i_{1,a}, r_{1,a}), (i_{2,a}, r_{2,a}), \dots, (i_{|I_a|,a}, r_{|I_a|,a})\}$, um conjunto de $|I_a|$ ¹ pares ordenados da forma $(i_{j,a}, r_{j,a})$ nos quais $i_{j,a}$ é uma imagem do produto anunciado em a e $r_{j,a}$ é um indicador quanto à presença de indícios de fraude em $i_{j,a}$. É importante se ressaltar que os rótulos das imagens podem ser independentes dos rótulos c presentes nos anúncios. Assim, um anúncio fraudulento pode conter todas as imagens com rótulo de não-fraude.

Denominada *Particionamento da Base de Dados*, a primeira etapa do método proposto consiste na separação de D em três conjuntos disjuntos, não vazios, treino (D_{tr}), validação (D_{val}) e teste (D_{tt}), tais que $D_{tr} \cup D_{val} \cup D_{tt} = D$. A partir dos percentuais

¹ $|I_a|$ indica a cardinalidade do conjunto de imagens presentes no anúncio a .

de anúncios a serem alocados em cada conjunto, esta etapa seleciona aleatoriamente os anúncios, assegurando que cada um dos três conjuntos apresente uma quantidade balanceada de fraudes e não fraudes. Em seguida, o conjunto teste é novamente separado em três conjuntos disjuntos, não vazios, treino (D'_{tr}), validação (D'_{val}) e teste (D'_{tt}), tais que $D'_{tr} \cup D'_{val} \cup D'_{tt} = D_{tt}$. Essa separação é importante para assegurar que a análise comparativa entre os resultados obtidos pelas diferentes alternativas de utilização do DFSC seja feita a partir dos dados de um mesmo conjunto (i.e., D'_{tt}). O detalhamento da utilização dessa separação será feito nas próximas seções.

Em seguida, na etapa denominada *Projeção de Dados*, cada um dos conjuntos D_X gerados pela etapa anterior (i.e., $X \in \{tr, val, tt\}$) é separado em dois subconjuntos, onde um deles contém somente dados estruturados $DE_X = \{(e_a, c_a)/a \in D_X\}$ e o outro possui somente imagens $DI_X = \bigcup_{a \in D_X} I_a$. Desta forma, D_{tr} , D_{val} e D_{tt} são separados em DE_{tr} , DI_{tr} , DE_{val} , DI_{val} , DE_{tt} e DI_{tt} . Cabe enfatizar que DE_{tt} é separado nos conjuntos DE'_{tr} , DE'_{val} e DE'_{tt} (i.e., $DE_{tt} = DE'_{tr} \cup DE'_{val} \cup DE'_{tt}$). De forma análoga, $DI_{tt} = DI'_{tr} \cup DI'_{val} \cup DI'_{tt}$.

4.1 DFSC(Est)

A Figura 12 apresenta uma ilustração gráfica das etapas específicas executadas pelo DFSC(Est).

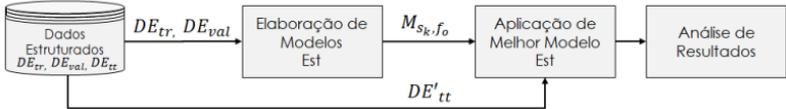


Figura 12 – DFSC(Est): Visão Geral das Etapas

A etapa *Elaboração de Modelos Est* constrói modelos de detecção de fraudes em anúncios, baseado exclusivamente em atributos estruturados de anúncios de *Social Commerce*. Para isso, utiliza os conjuntos DE_{tr} e DE_{val} e um conjunto de algoritmos de classificação $S = \{s_1, s_2, \dots, s_{|S|}\}$, em seguida os avalia conforme um conjunto de métricas de avaliação de desempenho $F = \{f_1, f_2, \dots, f_{|F|}\}$, sendo os dois últimos definidos pelo analista de dados, onde na ausência deste, seriam adotados conjuntos de algoritmos de aprendizado de máquina e métricas normalmente utilizados para problemas de classificação.

Assim, nesta etapa, cada algoritmo $s_k \in S$ é treinado a partir de um conjunto de valores de parâmetros ϕ de s_k e cada métrica de avaliação $f_o \in F$, gerando um modelo

de classificação de anúncios² m_{s_k} com os dados disponíveis em DE_{tr} , e os avalia com os dados em DE_{val} considerando f_o .

Este processo de treinamento se repete, a fim de identificar os N modelos distintos m_{s_k} que maximizem o desempenho de f_o em DE_{val} . Denominando tais modelos de $m_{s_k, f_o}^1, m_{s_k, f_o}^2, \dots, m_{s_k, f_o}^N$, esta etapa constrói o conjunto $M_{s_k, f_o} = \{m_{s_k, f_o}^1, m_{s_k, f_o}^2, \dots, m_{s_k, f_o}^N\}$, a ser utilizado nas próximas etapas.

Na etapa denominada *Aplicação de Melhores Modelo Est*, avalia os dados disponíveis em DE'_{tt} por cada elemento do conjunto M_{s_k, f_o} . Encaminhando os resultados obtidos para a próxima etapa.

Por fim, na etapa denominada *Análise de Resultados*, a partir dos desempenhos obtidos, esta etapa calcula a média e o desvio padrão de cada algoritmo e os armazena para fins de comparação com as demais alternativas de utilização do DFSC.

4.2 DFSC(Img)

A Figura 13 apresenta uma ilustração gráfica das etapas específicas executadas pelo DFSC(Img).

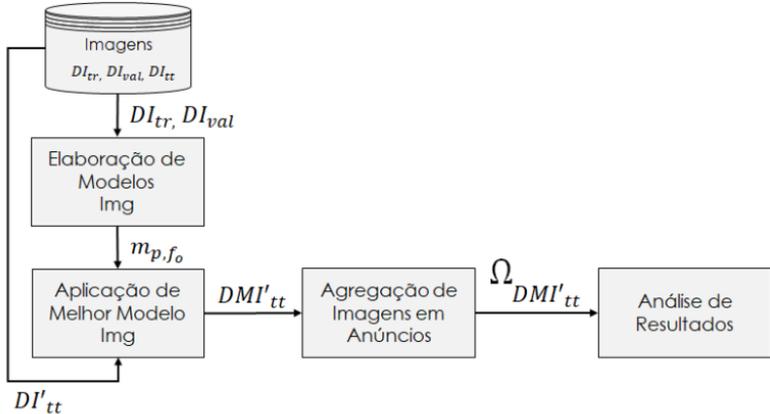


Figura 13 – DFSC(Img): Visão Geral das Etapas

A etapa *Elaboração de Modelos Img* constrói modelos de detecção de fraude,

²Embora os modelos $m_{s_k, \phi}$ sejam definidos em função do algoritmo de classificação s_k e pelo conjunto de parâmetros ϕ utilizados em sua geração, para fins de simplificação de notação, tais modelos serão representados apenas por m_{s_k} .

baseado exclusivamente em imagens de anúncios de *Social Commerce*. Para isso, utiliza os conjuntos DI_{tr} e DI_{val} , um conjunto de métricas de avaliação de desempenho F , conforme apresentado na seção 4.1, um algoritmo de aprendizado profundo³ p , sendo os dois últimos definidos pelo analista de dados. É importante ressaltar que, independente da arquitetura de aprendizado profundo escolhida, esta deve necessariamente conter uma camada *softmax* de saída. Tal requisito tem como objetivo assegurar que o modelo de classificação gerado seja capaz de indicar a probabilidade de que a imagem, a ele submetida, tenha indícios de fraude. Assim sendo, esta etapa treina um modelo de classificação de imagens m_{p,f_0} com os dados disponíveis em DI_{tr} , e o avalia com os dados em DI_{val} considerando f_0 . Este processo de treinamento se repete, a fim de identificar valores dos parâmetros de p que maximizem o desempenho de m_{p,f_0} em DI_{val} .

Conforme o próprio nome sugere, na etapa *Aplicação do Modelo Img*, o modelo m_{p,f_0} gerado na etapa anterior é aplicado sobre as imagens dos conjuntos DI_{tr} , DI_{val} e DI_{tt} , gerando os conjuntos DMI_{tr} , DMI_{val} e DMI_{tt} , respectivamente. Nesta etapa são obtidas as probabilidades de que as imagens contenham indícios de fraude. Assim, cada DMI_X é definido como $\{(i_{j,a}, r_{j,a}, m_{p,f_0}(i_{j,a})) / (i_{j,a}, r_{j,a}) \in DI_X\}$, onde $X \in \{tr, val, tt\}$ e $m_{p,f_0}(i_{j,a})$ indica a probabilidade da imagem $i_{j,a}$ conter indícios de fraude.

Na etapa *Agregação de Imagens em Anúncios*, o conjunto das probabilidades ($m_{p,f_0}(i_{j,a})$), geradas na etapa anterior associadas a cada imagem do anúncio, é definido por ω_a , onde $\omega_a = \{m_{p,f_0}(i_{1,a}), m_{p,f_0}(i_{2,a}), m_{p,f_0}(i_{3,a}), \dots, m_{p,f_0}(i_{|I_a|,a}) / i_{j,a} \in a\}$. A partir desde anúncio são extraídas diversas métricas, gerando para cada anúncio a o conjunto Ω_a . De forma que $\Omega_a = \{Min(\omega_a), Max(\omega_a), Q_1(\omega_a), Q_2(\omega_a), Q_3(\omega_a), \overline{\omega_a}, \sigma(\omega_a)\} / a \in DMI_X\}$. Desse modo, o conjunto Ω_a é utilizado para avaliação do anúncio permitindo que o conjunto seja usado total ou parcialmente.

Por fim, a etapa denominada *Análise de Resultados* tem como objetivo comparar as métricas escolhidas de Ω_a , no caso $\overline{\omega_a}$, comparadas com um valor de referência, geralmente 0.5, também chamado de *threshold*. Desse modo, nesta etapa são calculados a média e o desvio padrão, então estes valores são armazenados para fins de comparação com as demais alternativas de utilização do DFSC.

4.3 DFSC(Est+Img)

A Figura 14 apresenta uma ilustração gráfica das etapas específicas executadas pelo DFSC(Est+Img). A etapa *Elaboração de Modelos Img* constrói modelos de detecção de fraude, baseado exclusivamente em imagens de anúncios de *Social Commerce*. Para isso, utiliza os conjuntos DI_{tr} e DI_{val} , um conjunto de métricas de avaliação de desempenho F e

³Vide (36) para referencial teórico sobre algoritmos de aprendizado profundo.

um algoritmo de aprendizado profundo³ p , sendo os dois últimos definidos pelo analista de dados. É importante ressaltar que, independente da arquitetura de aprendizado profundo escolhida, esta deve necessariamente conter uma camada *softmax* de saída. Tal requisito tem como objetivo assegurar que o modelo de classificação gerado seja capaz de indicar a probabilidade de que a imagem a ele submetida tenha indícios de fraude. Esta etapa treina um modelo de classificação de imagens m_{p,f_o} com os dados disponíveis em DI_{tr} , e o avalia com os dados em DI_{val} considerando f_o . Este processo de treinamento se repete, a fim de identificar valores dos parâmetros de p que maximizem o desempenho de m_{p,f_o} em DI_{val} .

Conforme o próprio nome sugere, na etapa *Aplicação do Modelo Img*, o modelo m_{p,f_o} gerado na etapa anterior é aplicado sobre as imagens dos conjuntos DI_{tr} , DI_{val} e DI_{tt} , gerando os conjuntos DMI_{tr} , DMI_{val} e DMI_{tt} , respectivamente. Nesta etapa são obtidas as probabilidades de que as imagens contenham indícios de fraude. Assim, cada DMI_X é definido como $\{(i_{j,a}, r_{j,a}, m_{p,f_o}(i_{j,a})) / (i_{j,a}, r_{j,a}) \in DI_X\}$, onde $X \in \{tr, val, tt\}$ e $m_{p,f_o}(i_{j,a})$ indica a probabilidade da imagem $i_{j,a}$ conter indícios de fraude.

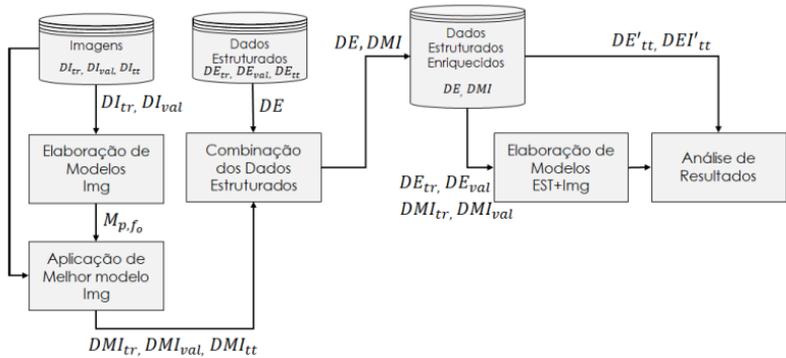


Figura 14 – DFSC(EST+Img): Visão Geral das Etapas

Denominada de *Combinação de Dados Estruturados*, a etapa seguinte tem como objetivo reunir as informações estruturadas dos anúncios com as probabilidades das imagens desses anúncios conterem indícios de fraude. Assim, os pares de conjuntos (DE_{tr}, DMI_{tr}) , (DE_{val}, DMI_{val}) e (DE_{tt}, DMI_{tt}) são combinados, gerando os conjuntos DEI_{tr} , DEI_{val} e DEI_{tt} respectivamente, onde $DEI_X = \{(e_a, c_a, \hat{\omega}_a, \tilde{\omega}_a) / (e_a, c_a) \in DE_X \wedge \hat{\omega}_a = \frac{\sum_{(i_{j,a}, r_{j,a}, m_{p,f_o}(i_{j,a})) \in DMI_X} m_{p,f_o}(i_{j,a})}{\tilde{\omega}_a} \wedge \tilde{\omega}_a = |\{(i_{j,a}, r_{j,a}, m_{p,f_o}(i_{j,a})) \in DMI_X\}|, X \in \{tr, val, tt\}$. Cabe destacar ainda que, de modo análogo ao particionamento de $DE_{tt} = DE'_{tr} \cup DE'_{val} \cup DE'_{tt}$ e $DI_{tt} = DI'_{tr} \cup DI'_{val} \cup DI'_{tt}$, DEI_{tt} também é particionado em

³Vide (36) para referencial teórico sobre algoritmos de aprendizado profundo.

$DEI'_{tr} \cup DEI'_{val} \cup DEI'_{tt}$. Lembrando, que essa subdivisão foi necessária para assegurar que o mesmo conjunto de dados seja avaliado por todas as alternativas de utilização DFSC, e será detalhado na próxima seção.

A etapa *Elaboração de Modelos Est+Img* tem como objetivo construir modelos de detecção de fraudes em anúncios a partir de dados estruturados e das probabilidades das imagens desses anúncios conterem indícios de fraude. Para tanto, recebe como entrada os conjuntos DEI_{tr} e DEI_{val} , além de um algoritmo de classificação s e uma métrica de avaliação de desempenho f , ambos previamente especificados pelo analista de dados. Assim, nesta etapa, o algoritmo s é treinado a partir de um conjunto de valores de parâmetros ϕ de s avaliado pela métrica f , gerando um modelo de classificação de anúncios⁴ m_s com os dados disponíveis em DEI_{tr} , e os avalia com os dados em DEI_{val} considerando f . Este processo de treinamento se repete, a fim de identificar os N modelos m_s com maior valor de f em DE_{val} .

Por fim, a etapa denominada *Análise de Resultados* tem como objetivo aplicar os N melhores modelos m_s identificados na etapa anterior sobre os dados disponíveis em DEI'_{tt} , avaliando-os por meio da métrica f . A partir dos N desempenhos obtidos, esta etapa calcula a média e o desvio padrão e os armazena para fins de comparação com as demais alternativas de utilização do DFSC.

4.4 DFSC(Est,Img)

A Figura 15 apresenta uma ilustração gráfica das etapas específicas executadas pelo DFSC(Est,Img).

⁴Embora o modelo $m_{s,\phi}$ seja definido em função do algoritmo de classificação s e pelo conjunto de parâmetros ϕ utilizados em sua geração, para fins de simplificação de notação, tal modelo será representado apenas por m_s .

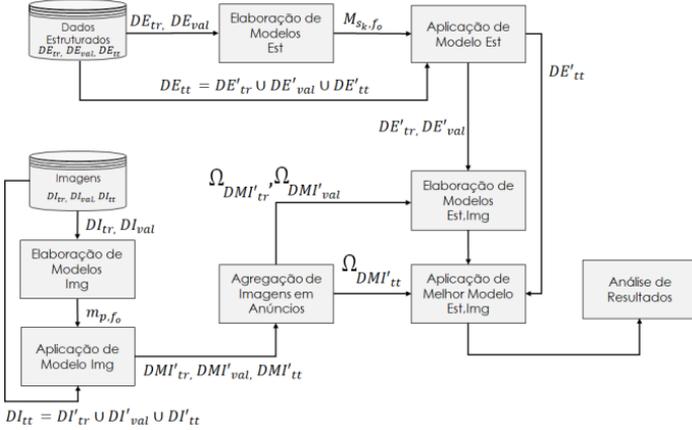


Figura 15 – DFSC(EST,Img): Visão Geral das Etapas

A etapa *Elaboração de Modelos Est* constrói modelos de detecção de fraudes em anúncios, baseado exclusivamente em atributos estruturados de anúncios de *Social Commerce*. Para isso, utiliza os conjuntos DE_{tr} e DE_{val} e um conjunto de algoritmos de classificação $S = \{s_1, s_2, \dots, s_{|S|}\}$, em seguida os avalia conforme um conjunto de métricas de avaliação de desempenho $F = \{f_1, f_2, \dots, f_{|F|}\}$, sendo os dois últimos definidos pelo analista de dados, onde na ausência deste, seriam adotados conjuntos de algoritmos de aprendizado de máquina e métricas normalmente utilizados para problemas de classificação.

Assim, nesta etapa, cada algoritmo $s_k \in S$ é treinado a partir de um conjunto de valores de parâmetros ϕ de s_k e cada métrica de avaliação $f_o \in F$, gerando um modelo de classificação de anúncios⁵ m_{s_k} com os dados disponíveis em DE_{tr} , e os avalia com os dados em DE_{val} considerando f_o .

Este processo de treinamento se repete, a fim de identificar os N modelos distintos m_{s_k} que maximizam o desempenho de f_o em DE_{val} . Denominando tais modelos de $m_{s_k, f_o}^1, m_{s_k, f_o}^2, \dots, m_{s_k, f_o}^N$, esta etapa constrói o conjunto $M_{s_k, f_o} = \{m_{s_k, f_o}^1, m_{s_k, f_o}^2, \dots, m_{s_k, f_o}^N\}$, a ser utilizado nas próximas etapas.

Na etapa denominada *Aplicação de Melhores Modelo Est*, são avaliados os dados disponíveis em $DE'_{tr}, DE'_{val}, DE'_{tt}$. De modo que nesta etapa, cada um dos modelos $m_{s_k, f_o}^z \in M_{s_k, f_o}$ a e_a , gerando uma primeira tupla de valores associados a a , $t_{est, a} = (m_{s_k, f_o}^1(e_a), m_{s_k, f_o}^2(e_a), \dots, m_{s_k, f_o}^N(e_a))$, onde $m_{s_k, f_o}^z(e_a)$ indica a classe (fraude, não fraude)

⁵Embora os modelos $m_{s_k, \phi}$ sejam definidos em função do algoritmo de classificação s_k e pelo conjunto de parâmetros ϕ utilizados em sua geração, para fins de simplificação de notação, tais modelos serão representados apenas por m_{s_k} .

predita por m_{s_k, f_o}^z baseado nos dados estruturados e_a do anúncio a .

A etapa *Elaboração de Modelos Img* constrói modelos de detecção de fraude, baseado exclusivamente em imagens de anúncios de *Social Commerce*. Para isso, utiliza os conjuntos DI_{tr} e DI_{val} , um conjunto de métricas de avaliação de desempenho F e um algoritmo de aprendizado profundo³ p , sendo os dois últimos definidos pelo analista de dados. É importante ressaltar que, independente da arquitetura de aprendizado profundo escolhida, esta deve necessariamente conter uma camada *softmax* de saída. Tal requisito tem como objetivo assegurar que o modelo de classificação gerado seja capaz de indicar a probabilidade de que a imagem a ele submetida tenha indícios de fraude. Esta etapa treina um modelo de classificação de imagens m_{p, f_o} com os dados disponíveis em DI_{tr} , e o avalia com os dados em DI_{val} considerando f_o . Este processo de treinamento se repete, a fim de identificar valores dos parâmetros de p que maximizem o desempenho de m_{p, f_o} em DI_{val} .

Conforme o próprio nome sugere, na etapa *Aplicação do Modelo Img*, o modelo m_{p, f_o} gerado na etapa anterior é aplicado sobre as imagens dos conjuntos DI_{tr} , DI_{val} e DI_{tt} , gerando os conjuntos DMI_{tr} , DMI_{val} e DMI_{tt} , respectivamente. Nesta etapa são obtidas as probabilidades de que as imagens contenham indícios de fraude. Assim, cada DMI_X é definido como $\{(i_{j,a}, r_{j,a}, m_{p, f_o}(i_{j,a})) / (i_{j,a}, r_{j,a}) \in DI_X\}$, onde $X \in \{tr, val, tt\}$ e $m_{p, f_o}(i_{j,a})$ indica a probabilidade da imagem $i_{j,a}$ conter indícios de fraude.

Na etapa *Agregação de Imagens em Anúncios*, o conjunto das probabilidades ($m_{p, f_o}(i_{j,a})$), geradas na etapa anterior associadas a cada imagem do anúncio, é definido por ω_a , onde $\omega_a = \{m_{p, f_o}(i_{1,a}), m_{p, f_o}(i_{2,a}), m_{p, f_o}(i_{3,a}), \dots, m_{p, f_o}(i_{|I_a|, a}) / i_{j,a} \in a\}$. A partir desde anúncio são extraídas diversas métricas, gerando para cada anúncio a o conjunto Ω_a . De forma que $\Omega_a = \{Min(\omega_a), Max(\omega_a), Q_1(\omega_a), Q_2(\omega_a), Q_3(\omega_a), \overline{\omega}_a, \sigma(\omega_a)\} / a \in DMI_X\}$. Desse modo, o conjunto Ω_a é representado no formato de tupla, compondo a segunda tupla de valores associada a a , de modo que $t_{img, a} = (Min(\omega_a), Max(\omega_a), Q_1(\omega_a), Q_2(\omega_a), Q_3(\omega_a), \overline{\omega}_a, \sigma(\omega_a))$. A tupla $t_{img, a}$ pode ter todos seus elementos utilizados ou apenas alguns deles.

A etapa *Elaboração de Modelos Est,Img* constrói modelos de detecção de fraude, baseado em atributos obtidos por meio da aplicação de outros modelos tanto em dados estruturados quanto em imagens, utilizando, desse modo, a probabilidade dos atributos estruturados conterem indícios de fraude das imagens desses anúncios conterem indícios de fraude. Para tanto recebe como entrada as tuplas $t_{est, a}$, $t_{img, a}$ onde $a \in DMI'_{tr} \cup DMI'_{val} \cup DE'_{tr} \cup DE'_{val}$. Em seguida, as duas tuplas $t_{est, a}$ e $t_{img, a}$ são concatenadas com o rótulo de a , gerando $h_a = (m_{s_k, f_o}^1(e_a), \dots, m_{s_k, f_o}^N(e_a), Min(\omega_a), \dots, \sigma(\omega_a), c_a)$.

Dessa forma, utilizando os conjuntos $h_a \in DE'_{tr} \cup DMI'_{tr} \cup DE'_{val} \cup DMI'_{val}$, e as métricas de avaliação de desempenho F , além do conjunto de algoritmos de classificação S ,

³Vide (36) para referencial teórico sobre algoritmos de aprendizado profundo.

onde cada algoritmo $\hat{s}_k \in S$ é treinado a partir de um conjunto de valores de parâmetros ϕ de \hat{s}_k e cada métrica de avaliação $f_o \in F$, gerando um modelo de classificação de anúncios⁶ $m_{\hat{s}_k}$ com os dados disponíveis em $DE'_{tr} \cup DMI'_{tr}$, e os avalia com os dados em $DE'_{val} \cup DMI'_{val}$ considerando f_o .

Este processo de treinamento se repete, a fim de identificar os N modelos distintos $\hat{m}_{\hat{s}_k}$ que maximizem o desempenho de f_o em $DE'_{val} \cup DMI'_{val}$. Denominando tais modelos de $\hat{m}_{\hat{s}_k, f_o}^1, \hat{m}_{\hat{s}_k, f_o}^2, \dots, \hat{m}_{\hat{s}_k, f_o}^N$, esta etapa constrói o conjunto $\hat{M}_{\hat{s}_k, f_o} = \{\hat{m}_{\hat{s}_k, f_o}^1, \hat{m}_{\hat{s}_k, f_o}^2, \dots, \hat{m}_{\hat{s}_k, f_o}^N\}$, a ser utilizado nas próximas etapas.

Por fim, a etapa denominada *Análise de Resultados* tem como objetivo aplicar os N melhores modelos \hat{m}_s identificados na etapa anterior sobre os dados disponíveis em $DE'_{tt} \cup DMI'_{tt}$, avaliando-os por meio da métrica f . A partir dos N desempenhos obtidos, esta etapa calcula a média e o desvio padrão e os armazena para fins de comparação com as demais alternativas de utilização do DFSC.

⁶Embora os modelos $m_{\hat{s}_k, \phi}$ sejam definidos em função do algoritmo de classificação \hat{s}_k e pelo conjunto de parâmetros ϕ utilizados em sua geração, para fins de simplificação de notação, tais modelos serão representados apenas por $m_{\hat{s}_k}$.

5 EXPERIMENTOS E RESULTADOS

Este capítulo descreve os experimentos realizados com as quatro alternativas de utilização do DFSC (DFSC(Est), DFSC(Img), DFSC(Est+Img) e DFSC(Est, Img)) e apresenta uma análise do desempenho de cada alternativa, além de análises comparativas de desempenho entre elas.

5.1 Protótipo

O protótipo utilizado para esta dissertação fora escrito em linguagem *Python 3.6.10*, utilizando as bibliotecas listadas no apêndice A. Além disto toda a arquitetura foi desenvolvida num ambiente na nuvem facilitando o tráfego de informações e incrementos ocasionais de poder computacional ou espaço para armazenamento da informação, que pode ser observado em detalhes na Tabela 9, bem como o tempo que cada instância foi utilizada. Para o treinamento de modelos de imagem as instâncias de computação acelerada, e para o treino de modelos estruturados se utilizou as instâncias otimizadas para computação, para as demais etapas do projeto foram utilizadas as outras instâncias.

Tabela 9 – Tempo computacional dos experimentos

GPU	vGPU	vCPU	Memória	Descrição	Tempo
0	0	2	4	Padrão	1.000,20
0	0	2	8	Padrão	490
0	0	4	16	Padrão	200
0	0	8	32	Padrão	7,07
0	0	8	16	Otimizadas para computação	235,82
1	16	8	61	Computação acelerada	0,50
4	64	32	244	Computação acelerada	2,01
Total de horas requeridas					1.935,59

5.2 Base de Dados Utilizada nos Experimentos

Foi realizada uma ampla busca por bases de dados sobre o tema desta dissertação e que contivesse simultaneamente dados estruturados sobre anúncios de produtos e imagens desses produtos. Entretanto, praticamente todas as bases de dados encontradas continham de forma mutuamente exclusiva, ou dados estruturados ou imagens, o que impossibilitava a realização de experimentos voltados à comprovação da hipótese de pesquisa levantada neste trabalho. Cabe ressaltar que (20), a única publicação que menciona a necessidade de se combinar dados estruturados de anúncios com imagens na detecção de fraudes, não disponibiliza a base de dados utilizada em sua pesquisa. Além disso, o sigilo dos dados foi

a uma alegação recorrente nos trabalhos relacionados para justificar a indisponibilidade dos dados das respectivas pesquisas. Tal fato justifica-se, uma vez que se tratam de dados sensíveis e de natureza classificada nos contextos responsáveis por sua coleta e armazenamento.

Diante do exposto, foi necessário estender a busca de forma a abranger empresas do mercado que atuassem em *Social Commerce* e que pudessem ter interesse na aplicação dos possíveis resultados que fossem gerados pelo presente trabalho. Uma das empresas consultadas mostrou interesse em ceder seus dados para uso nos experimentos.

Apesar da abertura, a empresa pediu sigilo tanto com relação aos dados que foram disponibilizados quanto ao seu nome, em função da Lei Geral de Proteção aos Dados (LGPD) prevista para entrar em vigor ao longo da execução desta pesquisa e diante da falta de clareza sobre como a lei iria abordar a sensibilidade dos conteúdos de imagem e campos de texto livre.

Ainda que não seja possível expor os dados disponibilizados pela empresa, para melhor elucidar ao leitor sobre algumas fraudes, foi seguido o tutorial do governo canadense (76), que ensina sobre as principais táticas usadas e o que deve ser observado a respeito do anúncio em questão:

- Avisos e reclamações postados online
- Ausência de telefone e *email* listado no anúncio / *website*
- Um nome diferente do nome da conta no cartão de crédito
- Ausência de rótulos na embalagem
- Baixa qualidade do produto
- Elevado desconto

Com base nisso, foi feita uma busca online em algumas plataformas de *social commerce* buscando alguns exemplos. Assim, serão apresentados alguns destes anúncios suspeitos, de modo anonimizado, para transferir o conhecimento ao leitor.

Ao observar o primeiro item, avisos e reclamações postados online, encontrou-se (77, 4) que trazem alguns exemplos textuais do que procurar ao olhar um anúncio e identificar fraudes. As imagens do site estão reproduzidas neste trabalho na Figura 16 e Figura 17. Em ambos os casos é observada uma expressão "pra rodar" que é uma expressão comum utilizada em serviços de venda online de carros sem documentação em dia ou com muita pendência financeira. Apesar de não ser ilegal a venda de carros com dívidas, se o usuário que está comprando esperar um produto sem pendências, entende-se que ocorreu uma fraude (classificada como 1.2, conforme Figura 6), uma vez que uma perda financeira

ocorre quando sua expectativa era outra. Contudo, caso o comprador tenha consciência das condições do produto e não tenha as intenções de legalizar e quitar todas as dívidas do carro, seria uma fraude (classificada como 2.1, conforme Figura 7), uma fraude contra a sociedade, porque as outras pessoas que detém carros tiveram de pagar as taxas ou multas que este usuário não irá pagar, bem como quando o usuário optar por fazer a interceptação de um carro roubado, sendo configurado não apenas como fraude mas outros crimes aos olhos da legislação brasileira.



Figura 16 – Exemplo 01 apresentado no site (4)

Na Figura 16, podem ser observados além do termo 'pra rodar', um elevado desconto (grátis), além da baixa qualidade da imagem (pode ter sido postada e feito captura de tela muitas vezes, cada transferência faz com que a qualidade da imagem reduza). Além disso, o ângulo da foto é pouco comum pra quem está planejando vender um carro, geralmente o foco é no carro, na posição horizontal.



Figura 17 – Exemplo 02 apresentado no site (4)

Na Figura 17, podem ser observados além do termo 'pra rodar', um elevado desconto (7.500 muito diferente da tabela FIPE para este veículo), na imagem é possível observar uma marca d'água no canto superior direito da imagem, além da descrição que tem a frase 'só para entendidos' o que torna o anúncio suspeito.

5.3 Pré-Processamento dos Dados

Os dados disponibilizados pela empresa continham informações sobre os anúncios, enriquecidos com informações dos usuários que fizeram a publicação. Parte das informações sobre os anúncios eram as imagens anunciadas. Como particularidade da base fornecida, os anúncios poderiam estar dispostos na base múltiplas vezes (isso ocorre devido a base fornecida ser transacional e à permissão de edição dos anúncios após sua publicação). Para o tratamento dessa múltipla ocorrência de mesmos anúncios ao longo do tempo, optou-se por ficar com a primeira inserção deste anúncio na plataforma apenas. Desse modo, usando como lógica que o ideal é que o anúncio fosse identificado como fraude no momento da sua inserção e não chegue sequer ao ponto de impactar algum outro usuário. Todo anúncio recebeu uma rotulagem - feita por especialistas de fraude da empresa - que classifica este entre fraude e não fraude. Além do anúncio, cada uma das imagens dos anúncios foi rotulada por um especialista entre Fraude e Não Fraude também, isso foi devido ao fato de que um anúncio pode ser fraude, porém não ter nenhuma imagem fraudulenta, bem como o inverso, o anúncio pode não ser fraude, apesar de ter pelo menos uma imagem fraudulenta. Desse modo a base não teve nenhum anúncio nenhuma entrada duplicada

bem como nenhum anúncio com duas rotulagens iguais ou distintas.

Apesar da anonimização feita pela empresa sobre os atributos disponibilizados, uma pesquisa em plataformas do setor automobilístico para fins comparativos, resultou na Tabela 10. Onde com o conhecimento de domínio é possível entender a importância do Ano, Marca, Modelo e Versão, pois estão diretamente relacionadas à uma tabela existente no mercado automobilístico chamada tabela FIPE, que permite ter preços comparativos do valor do carro em todo o mercado.

Tabela 10 – Tabela com exemplos de atributos utilizados por *e-commerce* e *social commerce*

Atributos	MercadoAvalia	Mobiauto	Webmotors	Facebook Marketplace	iCarros	Total
Ano do modelo	1	1	1	1	1	5
Marca	1	1	1		1	4
Modelo	1	1	1		1	4
Versão	1	1	1		1	4
Cor do Exterior	1	1	1	1		4
Ano de fabricação	1	1	1			3
Quilometragem	1	1		1		3
Blindado?		1	1			2
Portas	1	1				2
Câmbio	1	1				2
Combustível	1	1				2
Cor dos interiores				1		1
Placa	1					1
Preço				1		1
Renavam	1					1
Tipo de Transmissão				1		1
Valor	1					1
Total	13	11	7	6	4	41

Desse modo, com base na Tabela 10 é possível inferir que alguns dos campos que estão sendo avaliados pelo DFSC(Est) são: Ano do Modelo / Marca / Modelo / Versão além de dados como tipo câmbio (manual ou automático), número de portas, etc. Além desses, terão alguns atributos que serão combinações dos demais atributos que podem ser *booleanos* que sinalizam se um usuário preenheu ou não determinada informação, bem como se existe equivalência entre dados de duas ou mais colunas.

Na Tabela 11, é apresentada a distribuição dos registros únicos e distintos para cada classe, após o balanceamento dos dados utilizando a técnica de *downsample*.

Tabela 11 – Resumo Estatístico da Base de Dados

	Fraude	Não-Fraude
Anúncios	19.112	19.112
Imagens	43.066	34.234

Na primeira etapa do processamento do DFSC, foi realizado o particionamento da base de dados D nos conjuntos de treino (D_{tr}), validação (D_{val}) e teste (D_{tt}), aproximadamente na proporção 55%, 20%, 25%, respectivamente, levando às seguintes cardinalidades: $|D_{tr}| = 20.996$, $|D_{val}| = 7.530$ e $|D_{tt}| = 9.778$.

Como comentado no capítulo anterior, a fim de viabilizar a comparação das quatro alternativas de utilização do DFSC, é necessário que exista um conjunto de teste que será comum a todas as alternativas. Por conta disso, o conjunto D_{tt} foi subdividido em 3 grupos, gerando o conjunto, D' . Este particionamento seguiu a proporção de 80%, 10%, 10%, obtendo-se as seguintes cardinalidades: $|D'_{tr}| = 7822$, $|D'_{val}| = 978$ e $|D'_{tt}| = 978$. Desse modo, o conjunto $|D'_{tt}|$ será reservado para, ao final, permitir a comparação de desempenho entre as alternativas DFSC(Est), DFSC(Img), DFSC(Est+Img) e DFSC(Est, Img).

Já na etapa *Projeção de Dados*, cada um dos conjuntos gerados na etapa anterior foi separado em conjuntos de dados estruturados DE_X e conjuntos de imagens DI_X , onde $X \in \{tr, val, tt\}$. Detalhes estatísticos sobre cada conjunto podem ser observados na Tabela 12.

No total, cada tupla DE_X tinha 179 atributos, sendo um deles o rótulo do anúncio (c_a). Os atributos foram disponibilizados sendo 2 atributos textuais (*text_attribute001-002*), 11 atributos comportamentais (*behaviour_attribute001-011*), 73 atributos de anúncios (*ad_attribute001-073*), 101 atributos do usuário (*user_attribute001-101*), 1 campo data, 1 rótulo de classificação do anúncio, 1 campo com imagens, 2 campos de identificação, sendo um do usuário e um do anúncio para poder fazer o vínculo com o banco de dados de imagens.

Tabela 12 – Resumo Estatístico dos Conjuntos após a *Projeção de Dados*

Anúncios			Imagens		
Conjuntos	Fraude	Não-Fraude	Conjuntos	Fraude	Não-Fraude
DE_{tr}	10.458	10.458	DI_{tr}	22.921	18.531
DE_{val}	3.765	3.765	DI_{val}	8.445	6.654
DE_{tt}	4.889	4.889	DI_{tt}	11.700	9.049
DE'_{tr}	3911	3911	DI'_{tr}	9219	7130
DE'_{val}	489	489	DI'_{val}	1233	954
DE'_{tt}	489	489	DI'_{tt}	1248	965
Total	19.112	19.112		43.066	34.234

As próximas seções descrevem as especificidades dos experimentos e os resultados obtidos com cada alternativa de utilização do DFSC.

5.4 DFSC(Est)

Para a construção do DFSC(Est), foram utilizados os algoritmos de classificação K-Nearest-Neighbors (KNN)(78), Regressão Logística (LGR)(79), MultiLayer Perceptron (MLP)(80), Random Forest (RF)(81), Árvores de Decisão (DCT)(82), Análise de Discriminante Linear (LDA)(83), Naïve Bayes Gaussiano (GNB)(84), Support Vector Machine (SVM)(85).

5.4.1 Descrição

Os algoritmos selecionados estão expressos, na Tabela 13 por $S = \{RF, MLP, LDA, DCT, LGR, KNN, SVM, GNB\}$. Cada um destes foi avaliados conforme todas as combinações possíveis entre (Algoritmo, $\phi_{Parâmetros}$, $\phi_{Valores}$.)

Tabela 13 – Algoritmo e Parâmetros

Algoritmo	$\phi_{Parâmetros}$	$\phi_{Valores}$
RF	'n_estimators': 'min_samples_leaf': 'max_features':	[8, 16, 32, 64, 128, 256, 512, 1024, 2048, 3000], [1, 2, 5], [<i>'sqrt'</i> , 10, 11, 12, 14, 15, 16]
MLP	'solver': 'activation': 'learning_rate': 'alpha': 'hidden_layer_sizes': 'max_iter':	[<i>'lbfgs'</i> , <i>'sgd'</i> , <i>'adam'</i>], [<i>'identity'</i> , <i>'logistic'</i> , <i>'relu'</i> , <i>'tanh'</i>], [<i>'constant'</i> , <i>'invscaling'</i> , <i>'adaptive'</i>], [0.001], [3, 5, 10, 15, 20, 25, 30], [10000]
LDA	'solver': 'shrinkage':	[<i>'svd'</i> , <i>'lsqr'</i> , <i>'eigen'</i>], list(sorted(set(list(np.logspace(-3, 0, 7)) + list(np.logspace(-4, 0, 10)))))
DCT	'min_samples_leaf': 'max_features': 'splitter': 'criterion':	[1, 2, 5, 10, 15], [<i>'sqrt'</i> , 10, 11, 12, 14, 15, 16], [<i>'best'</i> , <i>'random'</i>], [<i>'gini'</i> , <i>'entropy'</i>]
LGR	'solver': 'penalty': 'C': 'max_iter':	[<i>'newton-cg'</i> , <i>'lbfgs'</i> , <i>'liblinear'</i> , <i>'sag'</i> , <i>'saga'</i>], [<i>'l1'</i> , <i>'l2'</i>], list(sorted(set(list(np.logspace(-3, 3, 7)) + list(np.logspace(-4, 4, 20))))) [100, 200, 4000]
KNN	'n_neighbors': 'leaf_size': 'algorithm': 'weights':	range(1, 30), [1, 2, 4, 8, 16, 30, 32, 64, 128, 256, 512], [<i>'auto'</i> , <i>'ball_tree'</i> , <i>'kd_tree'</i> , <i>'brute'</i>], [<i>'uniform'</i> , <i>'distance'</i>]
SVM	'C': 'gamma': 'kernel':	[0.01, 0.01, 0.1, 1, 10, 20, 50], [<i>'scale'</i> , <i>'auto'</i> , 0.001, 0.01, 0.1], [<i>'rbf'</i> , <i>'poly'</i> , <i>'sigmoid'</i> , <i>'linear'</i>]
GNB	-	-

Para este baseline foram utilizadas as métricas $F = \{acurácia (acc), precisão (prc), recall (rec), f1_score (fsc)\}$, e, para fins comparativos, será utilizada a métrica $f1_score$.

5.4.1.1 Critérios de Desempate

A escolha da métrica F-Score foi devido ao fato de que pondera precisão e recall numa única métrica, dessa forma sendo melhor para explicar melhorias, sua escolha também foi devido a algumas vulnerabilidades nas outras métricas que acabam desconsiderando os verdadeiros negativos, por vezes só avaliando se o algoritmo é bom em acertar amostras positivas, em caso de empate será seguida a ordem recall, acurácia, precisão como critérios de desempate. Todos os modelos foram treinados usando D_{tr} e avaliados comparativamente conforme D_{val} .

5.4.1.2 Melhores Modelos

A metodologia proposta requer que seja elegido pelo analista de dados um valor arbitrário (V_{arbt}) para definição do número de modelos por algoritmo. Para este experimento foi escolhido o $V_{arbt} = \text{cinco}$. A seguir, são selecionados os melhores modelos de modo que $V_{arbt} \leq N \leq 20 \iff (V_{arbt} \times |F|) \iff (5 \times 4)$.

O fato de N ser no máximo 20 é porque a arquitetura e modelo avaliados podem estar com desempenho melhor que os demais em mais de uma métrica e como ao final da seleção dos 20 melhores modelos, é aplicado um filtro de modelos distintos, logo a saída poderá ser menor que a entrada, porém será sempre maior ou igual ao V_{arbt} , uma vez que são escolhidas as melhores arquiteturas e não há duplicidade de arquitetura.

5.4.1.3 Melhor Modelo

Para seleção do melhor modelo, utilizam-se as métricas F conforme critérios de desempate, resultando num único modelo selecionado.

5.4.2 Resultado

Nesta seção são apresentados os resultados obtidos conforme a metodologia. Para a seleção das melhores parametrizações de cada um dos algoritmos, são comparados os resultados obtidos com cada um dos parâmetros no conjunto de validação D_{val} . Cada um desses modelos selecionados são então aplicados ao conjunto de teste, vale ressaltar que o melhor modelo é escolhido apenas com base na informação disponível na etapa de validação.

As parametrizações dos melhores modelos Est selecionados são apresentadas na Tabela 14, permitindo, um entendimento sobre quais geraram bons resultados. Desse modo na coluna algoritmo é apresentado o algoritmo que determinou os $\phi_{Parâmetros}$ a que este modelo está sujeito. Já na coluna, $\phi_{Valores}$, foram apresentados apenas os valores de cada um dos $\phi_{Parâmetros}$ dispostos seguindo a ordem da apresentação prévia na Tabela 13, para maior conforto do leitor, os $\phi_{Parâmetros}$ foram apresentados ao final de cada algoritmo na linha que apresenta a média e desvio padrão com relação aos diferentes $\phi_{Valores}$ estudados e selecionados. Com propósito informativos, a Tabela 14 traz o valor quando cada um de seus modelos, os selecionados conforme D_{val} , são aplicados ao conjunto teste D'_{tt} .

Tabela 14 – Melhores algoritmos DFSC(Est)

Algoritmo	ϕ_{desvio}	Conjuntos de Dados								
		D_{out}				D_{in}				
		Fraude F-Score	Acúrcia	Fraude Recall	Fraude Precisão	Fraude F-Score	Acúrcia	Fraude Recall	Fraude Precisão	
SVM	(50,auto,rbf)	0.801	0.772	0.917	0.801	0.710	0.733	0.672	0.900	0.618
SVM	(50,auto,rbf)	0.801	0.772	0.917	0.801	0.710	0.733	0.672	0.900	0.618
SVM	(20,0.01,rbf)	0.800	0.772	0.914	0.712	0.739	0.681	0.900	0.626	0.626
SVM	(10,0.01,rbf)	0.800	0.771	0.919	0.709	0.733	0.672	0.900	0.618	0.618
SVM	(50,0.01,rbf)	0.800	0.770	0.920	0.708	0.728	0.663	0.900	0.611	0.611
SVM	(20,auto,rbf)	0.800	0.770	0.919	0.708	0.740	0.675	0.925	0.617	0.617
SVM	(50,0.01,poly)	0.797	0.769	0.906	0.711	0.746	0.685	0.900	0.625	0.625
SVM	(50,auto,poly)	0.796	0.764	0.920	0.702	0.724	0.648	0.925	0.595	0.595
SVM	(20,0.01,poly)	0.796	0.766	0.911	0.707	0.724	0.648	0.925	0.595	0.595
SVM	(10,0.01,poly)	0.796	0.765	0.916	0.704	0.719	0.638	0.925	0.588	0.588
SVM	(10,auto,rbf)	0.796	0.764	0.918	0.702	0.740	0.675	0.925	0.617	0.617
$SVM\phi_{\text{desvio}} = \text{'solver', 'gamma', 'kernel'}$		0.798	0.768	0.916	0.707	0.733	0.666	0.915	0.611	0.611
Média ± Desvio Padrão		± 0.0022	± 0.0030	± 0.0046	± 0.0037	± 0.0087	± 0.0159	± 0.0129	± 0.0135	
MLP	(sgd,tanh,constant,0.001,30,10000)	0.796	0.765	0.917	0.703	0.764	0.706	0.950	0.639	0.639
MLP	(sgd,tanh,constant,0.001,30,10000)	0.796	0.765	0.917	0.703	0.764	0.706	0.950	0.639	0.639
MLP	(sgd,tanh,adaptive,0.001,30,10000)	0.795	0.765	0.915	0.703	0.764	0.706	0.950	0.639	0.639
MLP	(sgd,tanh,adaptive,0.001,20,10000)	0.792	0.759	0.919	0.696	0.740	0.675	0.925	0.617	0.617
MLP	(sgd,tanh,adaptive,0.001,10,10000)	0.791	0.759	0.913	0.698	0.758	0.697	0.950	0.631	0.631
MLP	(adam,relu,constant,0.001,15,10000)	0.790	0.764	0.887	0.712	0.718	0.666	0.850	0.621	0.621
MLP	(adam,relu,scaling,0.001,15,10000)	0.790	0.764	0.887	0.712	0.718	0.666	0.850	0.621	0.621
MLP	(adam,relu,adaptive,0.001,15,10000)	0.790	0.764	0.887	0.712	0.718	0.666	0.850	0.621	0.621
MLP	(sgd,tanh,constant,0.001,20,10000)	0.788	0.753	0.919	0.690	0.740	0.675	0.925	0.617	0.617
MLP	(sgd,relu,adaptive,0.001,25,10000)	0.787	0.752	0.916	0.690	0.764	0.706	0.950	0.639	0.639
MLP	(sgd,relu,constant,0.001,25,10000)	0.780	0.738	0.931	0.671	0.753	0.688	0.950	0.623	0.623
MLP $\phi_{\text{desvio}} = \text{'solver', 'activation', 'learning_rate', 'alpha', 'hidden_layer_size', 'max_iter'}$		0.790	0.758	0.909	0.699	0.744	0.685	0.915	0.627	0.627
Média ± Desvio Padrão		± 0.0045	± 0.0087	± 0.0157	± 0.0128	± 0.0109	± 0.0178	± 0.0459	± 0.0091	
RF	(512,1,15)	0.788	0.735	0.914	0.693	0.728	0.663	0.900	0.611	0.611
RF	(512,1,14)	0.788	0.734	0.914	0.692	0.717	0.644	0.900	0.596	0.596
RF	(566,1,14)	0.788	0.734	0.914	0.692	0.717	0.644	0.900	0.596	0.596
RF	(256,1,15)	0.788	0.733	0.915	0.691	0.722	0.654	0.900	0.603	0.603
RF	(1024,1,15)	0.787	0.733	0.914	0.691	0.717	0.644	0.900	0.596	0.596
RF	(16,1,15)	0.787	0.757	0.898	0.701	0.757	0.703	0.925	0.641	0.641
RF	(2048,1,15)	0.787	0.753	0.915	0.691	0.717	0.644	0.900	0.596	0.596
RF	(3000,1,16)	0.787	0.753	0.914	0.691	0.717	0.644	0.900	0.596	0.596
RF	(2048,1,16)	0.787	0.753	0.914	0.691	0.717	0.644	0.900	0.596	0.596
RF	(1024,1,16)	0.787	0.753	0.913	0.691	0.722	0.654	0.900	0.603	0.603
RF $\phi_{\text{desvio}} = \text{'n_estimators', 'min_samples_leaf', 'max_features'}$		0.787	0.754	0.913	0.693	0.723	0.654	0.903	0.603	0.603
Média ± Desvio Padrão		± 0.0004	± 0.0014	± 0.0051	± 0.0030	± 0.0125	± 0.0184	± 0.0079	± 0.0142	
KNN	(6,32,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,8,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,64,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,256,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,4,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,30,brute,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,256,ball,tree,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,256,auto,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,32,auto,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN	(6,256,kd,tree,uniform)	0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
KNN $\phi_{\text{desvio}} = \text{'n_neighbors', 'leaf_size', 'algorithm', 'weights'}$		0.777	0.748	0.878	0.696	0.718	0.666	0.850	0.621	0.621
Média ± Desvio Padrão		± 0.0001	± 0.0001	± 0.0001	± 0.0000	± 0.0000	± 0.0000	± 0.0000	± 0.0000	
LAGR	(lbgds,l2,6237767,39188721,100)	0.761	0.717	0.900	0.658	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,l2,100,0,100)	0.760	0.716	0.899	0.659	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,l2,100,0,200)	0.760	0.716	0.899	0.658	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,l2,345,5594781,6854,200)	0.760	0.716	0.898	0.659	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,l2,10,0,200)	0.760	0.716	0.899	0.658	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,l2,78,4759970351,4607,200)	0.760	0.716	0.899	0.658	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,linear,l2,1438,4498828766,4000)	0.760	0.716	0.898	0.659	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,linear,l2,1438,4498828766,100)	0.760	0.716	0.898	0.659	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,linear,l2,206,91380811149,4000)	0.760	0.716	0.898	0.658	0.753	0.688	0.950	0.623	0.623
LAGR	(lbgds,linear,l2,206,91380811149,100)	0.760	0.716	0.898	0.658	0.753	0.688	0.950	0.623	0.623
LAGR $\phi_{\text{desvio}} = \text{'solver', 'penalty', 'C', 'max_iter'}$		0.760	0.716	0.899	0.658	0.753	0.688	0.950	0.623	0.623
Média ± Desvio Padrão		± 0.0003	± 0.0002	± 0.0008	± 0.0002	± 0.0000	± 0.0000	± 0.0000	± 0.0000	
DCT	(5,14,best,gini)	0.756	0.720	0.871	0.669	0.688	0.625	0.825	0.590	0.590
DCT	(5,16,best,entropy)	0.756	0.720	0.865	0.670	0.739	0.681	0.900	0.626	0.626
DCT	(10,15,best,gini)	0.754	0.717	0.870	0.666	0.763	0.713	0.925	0.649	0.649
DCT	(5,15,best,gini)	0.753	0.715	0.867	0.665	0.746	0.685	0.925	0.625	0.625
DCT	(5,16,best,gini)	0.753	0.718	0.858	0.669	0.736	0.689	0.900	0.623	0.623
DCT	(10,14,best,entropy)	0.749	0.707	0.872	0.656	0.704	0.632	0.875	0.589	0.589
DCT	(10,15,best,entropy)	0.747	0.700	0.886	0.645	0.723	0.675	0.850	0.630	0.630
DCT	(10,14,best,gini)	0.747	0.699	0.886	0.645	0.704	0.632	0.875	0.589	0.589
DCT	(10,16,best,entropy)	0.747	0.700	0.883	0.647	0.704	0.632	0.875	0.589	0.589
DCT	(5,14,best,entropy)	0.746	0.704	0.873	0.652	0.712	0.635	0.900	0.588	0.588
DCT $\phi_{\text{desvio}} = \text{'min_samples_leaf', 'max_features', 'splitter', 'criterion'}$		0.751	0.710	0.874	0.659	0.747	0.669	0.825	0.590	0.590
Média ± Desvio Padrão		± 0.0040	± 0.0087	± 0.0093	± 0.0106	± 0.0256	± 0.0343	± 0.0316	± 0.0264	
LDA	(eigen,0.000774263682681127)	0.745	0.690	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.00074263682681127)	0.745	0.690	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.0002782559402207126)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.0002782559402207126)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(bnp,0.0001)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.0001)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.002154344690031882)	0.745	0.689	0.907	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.002154344690031882)	0.745	0.689	0.907	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(eigen,0.001)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA	(bnp,0.0001)	0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
LDA $\phi_{\text{desvio}} = \text{'solver', 'shrinkage'}$		0.745	0.689	0.906	0.632	0.714	0.629	0.925	0.581	0.581
Média ± Desvio Padrão		± 0.0001	± 0.0002	± 0.0005	± 0.0003	± 0.0000	± 0.0000	± 0.0000		

Então todos os modelos selecionados e apresentados em detalhe na Tabela 14 foram agrupados em relação a cada um dos algoritmos, permitindo a geração de métricas como média e desvio padrão por algoritmo. Dessa forma, os dados sumarizados são apresentados na Tabela 15. Cujas análises permitem observar que *GNB* não se adaptou bem a este problema, sendo o pior dentre todos os modelos. Contudo, para fins de pesquisa ele será mantido. Na tabela foram apresentadas cada uma das métricas em relação aos conjuntos de validação e teste utilizados.

Tabela 15 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para DFSC(Est) selecionados

Algoritmo	Conjuntos de Dados							
	D_{val}				D'_{tt}			
	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão
SVM	0,798 $\pm 0,0022$	0,768 $\pm 0,0030$	0,916 $\pm 0,0046$	0,707 $\pm 0,0037$	0,733 $\pm 0,0087$	0,666 $\pm 0,0159$	0,915 $\pm 0,0129$	0,611 $\pm 0,0135$
MLP	0,790 $\pm 0,0045$	0,758 $\pm 0,0087$	0,909 $\pm 0,0157$	0,699 $\pm 0,0128$	0,744 $\pm 0,0199$	0,685 $\pm 0,0178$	0,915 $\pm 0,0459$	0,627 $\pm 0,0091$
RF	0,787 $\pm 0,0004$	0,754 $\pm 0,0014$	0,913 $\pm 0,0051$	0,693 $\pm 0,0030$	0,723 $\pm 0,0125$	0,654 $\pm 0,0184$	0,903 $\pm 0,0079$	0,603 $\pm 0,0142$
KNN	0,777 $\pm 0,0001$	0,748 $\pm 0,0001$	0,878 $\pm 0,0001$	0,696 $\pm 0,0000$	0,718 $\pm 0,0000$	0,666 $\pm 0,0000$	0,850 $\pm 0,0000$	0,621 $\pm 0,0000$
LGR	0,760 $\pm 0,0003$	0,716 $\pm 0,0002$	0,899 $\pm 0,0008$	0,658 $\pm 0,0002$	0,753 $\pm 0,0000$	0,688 $\pm 0,0000$	0,950 $\pm 0,0000$	0,623 $\pm 0,0000$
DCT	0,751 $\pm 0,0040$	0,710 $\pm 0,0087$	0,873 $\pm 0,0093$	0,659 $\pm 0,0106$	0,724 $\pm 0,0256$	0,662 $\pm 0,0343$	0,885 $\pm 0,0316$	0,613 $\pm 0,0264$
LDA	0,745 $\pm 0,0001$	0,689 $\pm 0,0002$	0,906 $\pm 0,0005$	0,632 $\pm 0,0003$	0,714 $\pm 0,0000$	0,629 $\pm 0,0000$	0,925 $\pm 0,0000$	0,581 $\pm 0,0000$
GNB	0,578 $\pm 0,0000$	0,653 $\pm 0,0000$	0,475 $\pm 0,0000$	0,738 $\pm 0,0000$	0,563 $\pm 0,0000$	0,611 $\pm 0,0000$	0,500 $\pm 0,0000$	0,643 $\pm 0,0000$

5.5 DFSC(Img)

Para a construção do DFSC(Img), por ser destinada a manipular dados não estruturados, foi utilizada uma arquitetura de rede neural do tipo convolucional (do inglês, *Convolutional Neural Networks - CNN*), baseada em modelos com arquitetura *Residual* (86). No entanto, por não haver disponível uma base de imagens rotuladas com dados suficientes para modelos de aprendizado profundo, duas medidas foram tomadas para viabilizar a solução: foi utilizada a rede *MobileNetv2* (29) e incluiu-se a técnica de *Transfer Learning* (28). A arquitetura da rede *MobileNetv2* é conveniente por possuir menor número de camadas e, conseqüentemente, menor número de parâmetros a serem ajustados, o que se torna conveniente quando os bancos de dados de imagens são menores. Otimizando os esforços em se tirar o melhor proveito dos dados, a técnica de *Transfer Learning* concentra os esforços de aprendizado, para discriminação da imagem com e sem fraude, nas últimas camadas do modelo, sem atualização dos parâmetros das primeiras camadas, que, sendo uma evolução da *MobileNetv1*, com 28 camadas, treinadas em *TensorFlow* por meio da utilização de RMSprop, conforme apresentado (87, 88). A principal vantagem na adoção

da *MobileNet* é em função do menor uso de regularização e técnicas de geração de dados novos (*data augmentation*), como são usados diversos modelos menores, há um risco menor de *overfitting*.

5.5.1 Descrição

Para a escolha da arquitetura a ser utilizada, foram feitos testes de adições de camadas ao final do modelo, como tradicionalmente é feito em *Transfer Learning*, que podem ser observadas na Tabela 16.

Tabela 16 – Experimentação com alteração na quantidade de camadas

Número de camadas	F-score médio em D_{val}
1	32.3%
2	47.9%
3	59.3%
4	58.8%

A análise da Tabela 16 permite observar que o número de camadas 3 pareceu mais adequado para este experimento, principalmente ao levar em consideração o tempo requerido para cada uma das iterações. Cada uma das camadas adicionadas sempre continham 512 saídas, e a ultima camada sempre tendo apenas uma saída, desse modo foi possível comparar apenas a adição de camadas. A posteriori foram feitas experimentações em função das diversas funções de ativação como ReLu e Sigmóide, sendo apresentada a melhor arquitetura obtida na Figura 18, onde também pode-se observar o cuidado para minimização do risco de *overfit* por meio da inserção de camadas de *drop-out* entre elas.

```

Model: "sequential"
Layer (type)                Output Shape                Param #
-----
mobilenetv2_1.00_224 (Model) (None, 7, 7, 1280)        2257984
-----
global_average_pooling2d (G1 (None, 1280)                0
-----
dense (Dense)                (None, 512)                655872
-----
dropout (Dropout)            (None, 512)                0
-----
dense_3 (Dense)               (None, 512)                262656
-----
dropout_1 (Dropout)           (None, 512)                0
-----
dense_8 (Dense)               (None, 1)                  513
-----
Total params: 3,177,025
Trainable params: 919,041
Non-trainable params: 2,257,984

```

Figura 18 – Modelos Img - Arquitetura *CNN TransferLearning* (ImageNet adaptada)

5.5.1.1 Critérios de Desempate

Para escolha da melhor arquitetura, usou-se a métrica F-Score. Sua escolha foi devido ao fato de que pondera precisão e recall numa única métrica, dessa forma sendo melhor para explicar melhorias, sua escolha também foi devido a algumas vulnerabilidades nas outras métricas que acabam desconsiderando os verdadeiros negativos, por vezes só avaliando se o algoritmo é bom em acertar amostras positivas. Em caso de empate será seguida a ordem recall, acurácia, precisão como critérios de desempate. Todos os modelos foram treinados usando D_{tr} e avaliados comparativamente conforme D_{val} .

5.5.1.2 Melhores Modelos

Não se aplica a escolha dos melhores modelos para este momento.

5.5.1.3 Melhor Modelo

Uma vez escolhida a melhor arquitetura, para escolha do melhor modelo é analisada a *Loss Function* à tradução livre, função de perda. Onde, o momento de inflexão desta compreende o ponto mínimo do erro de validação, o que vai implicar no melhor momento da arquitetura, resultando num único modelo selecionado.

Conforme pode-se observar na Figura 19, para esta arquitetura escolhida, a melhor época de treinamento é entre 300 e 400, precisamente em 342, que será o modelo utilizado para futuras análises.

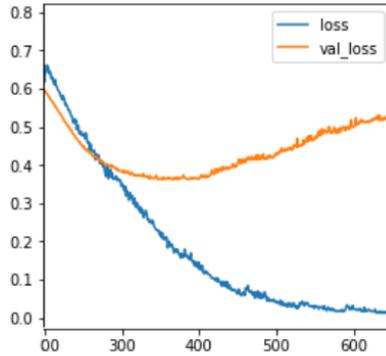
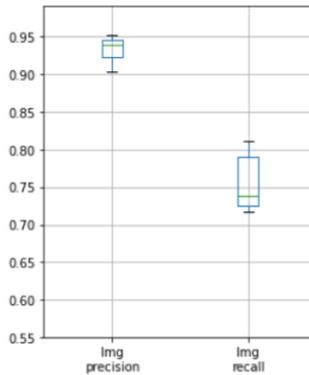


Figura 19 – Função de Perda da Melhor Arquitetura de Imagem

5.5.2 Resultado

Aplicar a arquitetura no conjunto DI_{val} , permitiu a escolha da melhor arquitetura e análise de resultado. Caso, esta abordagem seguisse a descoberta de melhores modelos desta arquitetura, estes podem ser observados na Figura 20. Cuja análise permite afirmar que analisar apenas a imagem é possível com uma alta precisão (com baixo desvio padrão) e um *recall* próximo de 0.75, porém com um desvio padrão maior, considerando que a cada anúncio as imagens serão agregadas conforme Ω_a e a métrica que foi avaliada foi a média, representada por $\overline{\omega_a}$, desse modo, se o valor fosse maior ou igual a 0.5, este anúncio seria considerado Fraude.

Figura 20 – Aplicação da melhor arquitetura obtida por DFSC(Img) ao conjunto DI_{val}

Além disso, é possível observar a performance do modelo quando aplicado ao Conjunto de Teste D_{tt} , na Tabela 17.

Tabela 17 – FDSC(Img) melhor arquitetura e melhor modelo aplicado ao conjunto de Teste D_{tt}

	F-Score	Acurácia	Recall	Precisão
FDSC(Img)	0,719	0,809	0,575	0,958

Apesar de não ser indispensável à pesquisa proposta, esta etapa foi realizada para que a pesquisa também seja capaz de responder a pergunta *"mas usar apenas a imagem, não seria, por si só, suficiente à melhora no resultado?"*, é importante destacar que esta abordagem, apesar do baixo recall, menos de 60%, apresenta uma altíssima precisão, acima de 95%, o que pode sugerir que existe algumas poucas imagens que tem características de fraude muito evidentes que o modelo de imagem consegue captar e interpretar este sinal corretamente.

5.6 Critérios para verificação de relevância estatística

Para verificar a relevância estatística dos resultados obtidos, aplicou-se o Teste-t pareado, com o objetivo de verificar se a média das diferenças entre duas amostras pareadas é diferente de 0. Este teste foi escolhido, pois sua utilização é indicada quando se deseja comparar o mesmo conjunto de itens após a medição sob duas condições diferentes, no caso, DFSC(Est) comparado com DFSC(Est+Img) ou DFSC(Est,Img), conforme serão apresentados nas seguintes subseções. Cabe destacar ainda que o Teste-t pareado requer:

- Distribuição dos dados seja aproximadamente uma distribuição Normal. Ainda que a distribuição não seja Normal (desde que a variação seja pequena), o Teste-T pareado é robusto à tal violação, sendo considerado aplicável nessas circunstâncias.
- Não existam *outliers* significativos.

Caso o Teste-t pareado não seja aplicado, então fez-se a aplicação do teste Wilcoxon.

5.7 DFSC(Est+Img)

5.7.1 Descrição

Nesta seção encontra-se descrita a configuração utilizada nos experimentos com o método DFSC(Est+Img).

A arquitetura de aprendizado profundo e a parametrização utilizadas nos experimentos com a DFSC(Img) foram adotadas na etapa *Elaboração de Modelos Img*.

De forma análoga aos experimentos realizados nas demais implementações do método DFSC na etapa de construção dos modelos baseados em dados estruturados, os algoritmos de classificação utilizados na etapa *Elaboração de Modelos EST+Img* foram o SVM, MLP, Random Forest, K-NN, LCR, DCT, LDA e GNB. Acurácia, F-Score, Precisão e *recall* foram as métricas de avaliação empregadas também nesta etapa. A quantidade de melhores modelos considerados foi $N = 10$. DEI_{tr} e DEI_{val} foram os conjuntos de dados utilizados na referida etapa.

Por fim, na etapa *Análise de Resultados*, o conjunto de dados utilizado para fins de avaliação dos modelos foi o DEI'_{tt} .

5.7.2 Resultado

As Tabelas 18 e 19 apresentam os resultados obtidos com o DFSC(Est+Img). Enquanto a primeira tabela mostra os desempenhos dos 10 melhores modelos gerados a partir de cada algoritmo de classificação, a segunda resume a média e o desvio padrão dos desempenhos desses modelos no conjunto de testes (DEI'_{tt}). Na Tabela 18, foram apresentados na coluna algoritmo é apresentado o algoritmo que determinou os $\phi_{Parâmetros}$ a que este modelo está sujeito. Já na coluna, $\phi_{Valores}$, foram apresentados apenas os valores de cada um dos $\phi_{Parâmetros}$ dispostos seguindo a ordem da apresentação prévia na Tabela 13, para maior conforto do leitor, os $\phi_{Parâmetros}$ foram apresentados ao final de cada algoritmo na linha que apresenta a média e desvio padrão com relação aos diferentes $\phi_{Valores}$ estudados e selecionados. Com propósito informativos, a Tabela 18 traz o valor quando cada um de seus modelos, os selecionados conforme DEI_{val} , são aplicados ao conjunto teste DEI'_{tt} .

Então todos os modelos selecionados e apresentados em detalhe na Tabela 18 foram agrupados em relação a cada um dos algoritmos, permitindo a geração de métricas como média e desvio padrão por algoritmo. Dessa forma, os dados sumarizados são apresentados na Tabela 19. Cujas análises permitem observar que *GNB* não se adaptou bem a este problema, sendo o pior dentre todos os modelos. Contudo, para fins de pesquisa ele será mantido. Na tabela foram apresentadas cada uma das métricas em relação aos conjuntos de validação e teste utilizados.

Tabela 19 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para Melhores algoritmos DFSC(Est+Img)

Algoritmo Média \pm Desvio Padrão	Conjuntos de Dados							
	D_{val}				D_{μ}			
	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão
SVM	0,813 $\pm 0,0022$	0,789 $\pm 0,0030$	0,917 $\pm 0,0046$	0,730 $\pm 0,0037$	0,753 $\pm 0,0087$	0,704 $\pm 0,0159$	0,903 $\pm 0,0129$	0,646 $\pm 0,0135$
MLP	0,806 $\pm 0,0045$	0,780 $\pm 0,0087$	0,913 $\pm 0,0157$	0,722 $\pm 0,0128$	0,745 $\pm 0,0199$	0,692 $\pm 0,0178$	0,900 $\pm 0,0459$	0,635 $\pm 0,0091$
RF	0,797 $\pm 0,0004$	0,766 $\pm 0,0014$	0,917 $\pm 0,0051$	0,704 $\pm 0,0030$	0,762 $\pm 0,0125$	0,709 $\pm 0,0184$	0,930 $\pm 0,0079$	0,645 $\pm 0,0142$
KNN	0,777 $\pm 0,0001$	0,750 $\pm 0,0001$	0,872 $\pm 0,0001$	0,701 $\pm 0,0000$	0,721 $\pm 0,0000$	0,681 $\pm 0,0000$	0,825 $\pm 0,0000$	0,641 $\pm 0,0000$
LGR	0,775 $\pm 0,0003$	0,738 $\pm 0,0002$	0,905 $\pm 0,0008$	0,678 $\pm 0,0002$	0,764 $\pm 0,0000$	0,706 $\pm 0,0000$	0,950 $\pm 0,0000$	0,639 $\pm 0,0000$
DCT	0,759 $\pm 0,0040$	0,725 $\pm 0,0087$	0,866 $\pm 0,0093$	0,676 $\pm 0,0106$	0,734 $\pm 0,0256$	0,683 $\pm 0,0343$	0,873 $\pm 0,0316$	0,634 $\pm 0,0264$
LDA	0,758 $\pm 0,0001$	0,708 $\pm 0,0002$	0,916 $\pm 0,0005$	0,647 $\pm 0,0003$	0,765 $\pm 0,0000$	0,700 $\pm 0,0000$	0,975 $\pm 0,0000$	0,629 $\pm 0,0000$
GNB	0,581 $\pm 0,0000$	0,655 $\pm 0,0000$	0,477 $\pm 0,0000$	0,742 $\pm 0,0000$	0,568 $\pm 0,0000$	0,620 $\pm 0,0000$	0,500 $\pm 0,0000$	0,659 $\pm 0,0000$

5.7.3 Análise de Resultados

Os resultados dos experimentos com os métodos DFSC(Est) e DFSC(Est+Img) estão sumarizados na Tabela 20. Nela, cada célula indica a média (μ) \pm desvio padrão (σ) da métrica correspondente, decorrente da aplicação dos 10 melhores modelos aplicados no conjunto DEI'_t . O melhor resultado obtido a cada métrica e algoritmo encontra-se destacado em negrito. A tabela também permitiu observar que a inclusão da imagem apresentou ganhos em todos os algoritmos quando comparados com DFSC(Est).

Em uma primeira análise, pode-se observar que os valores das médias de F-Score, Acurácia e Precisão obtidas com o DFSC(Est+Img) foram superiores aos valores correspondentes obtidos com o DFSC(Est). Tal fato se constituiu em evidência experimental que aponta para a validade da hipótese levantada por este trabalho de que considerar imagens sobre os produtos de forma combinada com os dados estruturados de anúncios em *social commerce* pode melhorar o desempenho dos métodos de detecção de fraudes que se baseiam exclusivamente nos dados estruturados desses anúncios.

Tabela 20 – Tabela Comparativa de Médias \pm Desvio Padrão ($\mu \pm \sigma$) para Melhores Modelos dos Métodos DFSC(Est) e DFSC(Est+Img) aplicados ao conjunto Teste (DEI_{tt})

Algoritmos	F-Score		Acurácia		Recall		Precisão	
	DFSC (Est)	DFSC (Est+Img)						
LDA	0.714 ± 0.0000	0.765 ± 0.0000	0.629 ± 0.0000	0.700 ± 0.0000	0.925 ± 0.0000	0.975 ± 0.0000	0.581 ± 0.0000	0.629 ± 0.0000
LGR	0.753 ± 0.0000	0.764 ± 0.0000	0.688 ± 0.0000	0.706 ± 0.0000	0.950 ± 0.0000	0.950 ± 0.0000	0.623 ± 0.0000	0.639 ± 0.0000
RF	0.723 ± 0.0125	0.762 ± 0.0094	0.654 ± 0.0184	0.709 ± 0.0121	0.903 ± 0.0079	0.930 ± 0.0105	0.603 ± 0.0142	0.645 ± 0.0090
SVM	0.733 ± 0.0087	0.753 ± 0.0111	0.666 ± 0.0159	0.704 ± 0.0157	0.915 ± 0.0129	0.903 ± 0.0142	0.611 ± 0.0135	0.646 ± 0.0138
MLP	0.744 ± 0.0199	0.745 ± 0.0118	0.685 ± 0.0178	0.692 ± 0.0148	0.915 ± 0.0459	0.900 ± 0.0167	0.627 ± 0.0091	0.635 ± 0.0113
DCT	0.724 ± 0.0256	0.734 ± 0.0234	0.662 ± 0.0343	0.683 ± 0.0321	0.885 ± 0.0316	0.872 ± 0.0381	0.613 ± 0.0264	0.634 ± 0.0282
KNN	0.718 ± 0.0000	0.721 ± 0.0000	0.666 ± 0.0000	0.681 ± 0.0000	0.850 ± 0.0000	0.825 ± 0.0000	0.621 ± 0.0000	0.641 ± 0.0000
GNB	0.562 ± 0.0000	0.568 ± 0.0000	0.611 ± 0.0000	0.620 ± 0.0000	0.500 ± 0.0000	0.500 ± 0.0000	0.643 ± 0.0000	0.659 ± 0.0000

As próximas subseções apresentam uma análise detalhada ao utilizar DFSC(Est) ou DFSC(Est+Img) e sua significância estatística (obtida por meio da utilização dos conceitos descritos em 5.6) dos resultados considerando cada algoritmo de classificação e cada métrica de avaliação separadamente à exceção de GNB que não pode ter sua contribuição estatística avaliada por ter tido apenas uma instância.

5.7.3.1 Avaliações e relevância estatística de cada uma das métricas

Para confirmar o aparente ganho é necessário que isto seja validado estatisticamente, desse modo, na Tabela 21, foram apresentadas, para cada algoritmo a avaliação individual de cada métrica, qual foi o teste aplicado para descobrir normalidade e seus respectivos valores. Além da conclusão se a aplicação do teste de normalidade concluiu ser uma distribuição normal ou não. A partir disso foram aplicados os Testes de Hipótese, como era condicional a cada tipo de distribuição eles são explicitados e seus respectivos valores registrados. Por fim, a última coluna registra se é possível afirmar que a inclusão da imagem exerceu influencia estatisticamente significativa ou se seria necessário descartar essa hipótese podendo acontecer ao acaso. Maiores detalhes com relação a esta avaliação estatística poderá ser encontrada no apêndice B. Desse modo, uma análise permite observar que quase todas as métricas de todos os modelos tiveram diferença significativa. E que apenas os modelos MLP e DCT não puderam registrar alterações significativas em suas métricas F-Score, reforçando a hipótese inicial de que considerar as imagens para os modelos de detecção em *social commerce* contribuiria para a qualidade da detecção.

Tabela 21 – Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem

Algoritmo	Métrica Avaliada	Teste de Normalidade (Shapiro-Wilk)	Distribuição Normal?	Teste de Hipótese	Valores no Teste de Hipótese	Estatisticamente significativa?
LDA	F-Score	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LDA	Acurácia	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LDA	Recall	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LDA	Precisão	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LGR	F-Score	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LGR	Acurácia	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
LGR	Recall	(1.00, 1.000)	Sim	T-pareado	(nan, nan)	Não
LGR	Precisão	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
RF	F-Score	(0.77, 0.003)	Não	Wilcoxon	(2.00, 0.002)	Sim
RF	Acurácia	(0.78, 0.004)	Não	Wilcoxon	(2.00, 0.002)	Sim
RF	Recall	(0.72, 0.001)	Não	Wilcoxon	(5.00, 0.003)	Sim
RF	Precisão	(0.78, 0.004)	Não	Wilcoxon	(2.00, 0.002)	Sim
SVM	F-Score	(0.86, 0.054)	Sim	T-pareado	(-5.19, 0.000)	Sim
SVM	Acurácia	(0.93, 0.346)	Sim	T-pareado	(-6.65, 0.000)	Sim
SVM	Recall	(0.59, 0.000)	Não	Wilcoxon	(0.00, 0.102)	Não
SVM	Precisão	(0.96, 0.815)	Sim	T-pareado	(-7.63, 0.000)	Sim
MLP	F-Score	(0.85, 0.055)	Sim	T-pareado	(-0.17, 0.871)	Não
MLP	Acurácia	(0.83, 0.035)	Não	Wilcoxon	(18.00, 0.326)	Não
MLP	Recall	(0.71, 0.001)	Não	Wilcoxon	(24.00, 0.717)	Não
MLP	Precisão	(0.86, 0.084)	Sim	T-pareado	(-1.69, 0.126)	Não
DCT	F-Score	(0.96, 0.699)	Sim	T-pareado	(-1.55, 0.146)	Não
DCT	Acurácia	(0.95, 0.535)	Sim	T-pareado	(-2.31, 0.038)	Sim
DCT	Recall	(0.95, 0.505)	Sim	T-pareado	(0.47, 0.645)	Não
DCT	Precisão	(0.93, 0.302)	Sim	T-pareado	(-2.79, 0.015)	Sim
KNN	F-Score	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
KNN	Acurácia	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim
KNN	Recall	(1.00, 1.000)	Sim	T-pareado	(inf, 0.000)	Sim
KNN	Precisão	(1.00, 1.000)	Sim	T-pareado	(-inf, 0.000)	Sim

5.8 DFSC(Est,Img)

5.8.1 Descrição

Nesta seção encontra-se descrita a configuração utilizada nos experimentos com o método DFSC(Est,Img)¹.

Em resumo, neste experimento, são combinadas as saídas de DFSC(Est) com DFSC(Img), então ambas são avaliadas por um algoritmo. Antes de detalhar as abordagens,

¹No apêndice D foram feitas outras investigações com relação a esta abordagem

é importante reforçar que, nesta não foi feita a avaliação do algoritmo SVM, uma vez que sua saída não é probabilística, e a abordagem sugerida pelo *scikit-learn* iria introduzir uma nova complexidade sob a qual não se poderia manter o controle do experimento.

A arquitetura de aprendizado profundo e a parametrização utilizadas nos experimentos com a DFSC(Img) foram adotadas na etapa *Elaboração de Modelos Img*. Análogamente, na etapa *Elaboração de Modelos Est*, foram utilizadas as arquiteturas e parametrizações de DFSC(Est). A quantidade de modelos considerados foi $N = 10$.

Desse modo, primeiro aplica-se o DFSC(Img), gerando para cada anúncio $\overline{\omega}_a$ e $|\omega_a|$. Em seguida, obtém-se um conjunto de resultados da porção DFSC(Est), agregado com os resultados de DFSC(Img) e, por meio de um algoritmo, nesta seção chamado de *combinador*, classificou o anúncio, o que configura o DFSC(Est,Img). Nesta experiência fez-se necessária a utilização dos conjuntos $DE'_{tr}, DE'_{val}, DMI'_{tr}, DMI'_{val}$, pois era necessário que o modelo a ser treinado na etapa *Elaboração de Modelos Est,Img* nunca tivesse sido exposto aos dados previamente para evitar contaminação.

De forma análoga aos experimentos realizados nas demais implementações do método DFSC na etapa de construção dos modelos baseados em dados estruturados, os algoritmos de classificação utilizados na etapa *Elaboração de Modelos Est,Img* foram o MLP, Random Forest, K-NN, LCR, DCT, LDA e GNB. Acurácia, F-Score, Precisão e *recall* foram as métricas de avaliação empregadas também nesta etapa. A quantidade de melhores modelos considerados foi $N = 10$. $DE'_{tr}, DE'_{val}, DMI'_{tr}$ e DMI'_{val} foram os conjuntos de dados utilizados na referida etapa. Deste modo, nesta organização dos dados são gerados os melhores modelos de dados estruturados (em detalhe, é como se tivesse uma tabela, onde cada linha representa uma entidade avaliada, cada coluna um modelo avaliador e as colunas são agrupadas em função do algoritmo principal, o que permite obter tanto o melhor por grupo/algoritmo quanto o melhor global/de todas as colunas), bem como o melhor modelo de dado de imagem (em detalhe, uma coluna com a classificação média das imagens daquele anúncio). Dessa forma algumas colunas são selecionadas e submetidas novamente a um algoritmo classificador (ou seja, são escolhidas algumas colunas que podem ser o melhor de todos os estruturados, com o melhor de imagem; o melhor de cada um dos estruturados com o melhor de imagem), o que permite que sejam feitos vários experimentos para decidir a melhor forma de combinar estes atributos.

Por fim, na etapa *Análise de Resultados*, os conjuntos de dados utilizado para fins de avaliação dos modelos foi o DE'_u e DMI'_u . Nota-se que para garantir a comparabilidade da inserção da imagem ou não, neste método foram executados os mesmos processos, foi submetido a um combinador um conjunto sem atributos de imagem e um conjunto com atributos de imagem para então ser contabilizada sua avaliação. Ao conjunto sem imagem foi atribuído a nomenclatura de DFSC(Est) e ao conjunto com imagem DFSC(Est,Img).

5.8.2 Resultado

A utilização do Combinador de DFSC(Est,Img), permitiu que os resultados pudessem ser resumidos por média e desvio padrão dos 10 melhores modelos gerados na etapa *Elaboração de Modelos Est,Img*, aplicados tanto ao conjunto de validação (DE'_{val} , DMI'_{val}) quanto testes (DE'_{tt} , DMI'_{tt}). Na Tabela 22 foram apresentados os modelos combinadores, cada uma das métricas escolhidas, além do desvio padrão e média ao considerar determinado algoritmo como algoritmo combinador. A partir da tabela, é observado que, aparentemente, nenhum modelo combinador seria descartado.

Tabela 22 – Médias \pm Desvio Padrão ($\mu \pm \sigma$) obtido para Melhores algoritmos DFSC(Est,Img)

Combinador Média \pm Desvio Padrão	Conjuntos de Dados							
	D_{val}				D_{tt}			
	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão	Fraude F-Score	Acurácia	Fraude Recall	Fraude Precisão
LDA	0.910 ± 0.0299	0.912 ± 0.0262	0.896 ± 0.0598	0.926 ± 0.0153	0.879 ± 0.0236	0.883 ± 0.0207	0.852 ± 0.0422	0.910 ± 0.0181
LGR	0.917 ± 0.0269	0.918 ± 0.0240	0.914 ± 0.0480	0.922 ± 0.0083	0.884 ± 0.0229	0.886 ± 0.0202	0.870 ± 0.0414	0.899 ± 0.0141
RF	0.908 ± 0.0212	0.908 ± 0.0204	0.904 ± 0.0326	0.913 ± 0.0216	0.867 ± 0.0257	0.869 ± 0.0240	0.858 ± 0.0405	0.877 ± 0.0233
MLP	0.910 ± 0.0264	0.911 ± 0.0238	0.906 ± 0.0474	0.916 ± 0.0131	0.890 ± 0.0237	0.891 ± 0.0211	0.886 ± 0.0464	0.895 ± 0.0175
DCT	0.862 ± 0.0349	0.863 ± 0.0345	0.857 ± 0.0459	0.869 ± 0.0427	0.825 ± 0.0313	0.830 ± 0.0287	0.802 ± 0.0457	0.850 ± 0.0309
KNN	0.881 ± 0.0305	0.882 ± 0.0284	0.883 ± 0.0504	0.881 ± 0.0270	0.850 ± 0.0431	0.852 ± 0.0408	0.841 ± 0.0564	0.860 ± 0.0383
GAU	0.809 ± 0.0499	0.818 ± 0.0247	0.806 ± 0.1658	0.854 ± 0.0938	0.787 ± 0.0454	0.803 ± 0.0240	0.754 ± 0.1500	0.868 ± 0.1055

5.8.3 Análise de Resultados

Os resultados dos experimentos com os métodos DFSC(Est) e DFSC(Est,Img) estão sumarizados na Tabela 23. Nela, cada célula indica a média (μ) \pm desvio padrão (σ) da métrica correspondente, decorrente da aplicação dos 10 melhores modelos aplicados ao conjunto de testes (DE'_{tt} , DMI'_{tt}). A tabela também permitiu observar que a inclusão da imagem apresentou ganhos em todos os algoritmos e todas as métricas quando comparados com DFSC(Est) à exceção de GNB na métrica de Recall.

Tabela 23 – Comparação DFSC(Est) e DFSC(Est,Img) de Médias \pm Desvio Padrão ($\mu \pm \sigma$) aplicados ao conjunto Teste (DEI_t)

Algoritmos	F-Score		Acurácia		Recall		Precisão	
	DFSC (Est)	DFSC (Est,Img)						
LDA	0.768 ± 0.0315	0.879 ± 0.0236	0.762 ± 0.0298	0.883 ± 0.0207	0.787 ± 0.0470	0.852 ± 0.0422	0.750 ± 0.0285	0.910 ± 0.0181
LGR	0.749 ± 0.0371	0.884 ± 0.0229	0.756 ± 0.0339	0.886 ± 0.0202	0.731 ± 0.0530	0.870 ± 0.0414	0.770 ± 0.0362	0.899 ± 0.0141
RF	0.721 ± 0.0475	0.867 ± 0.0257	0.741 ± 0.0390	0.869 ± 0.0240	0.674 ± 0.0618	0.858 ± 0.0405	0.778 ± 0.0394	0.877 ± 0.0233
MLP	0.750 ± 0.0377	0.890 ± 0.0237	0.761 ± 0.0339	0.891 ± 0.0211	0.722 ± 0.0539	0.886 ± 0.0464	0.784 ± 0.0384	0.895 ± 0.0175
DCT	0.686 ± 0.0470	0.825 ± 0.0313	0.704 ± 0.0397	0.830 ± 0.0287	0.647 ± 0.0626	0.802 ± 0.0457	0.732 ± 0.0459	0.850 ± 0.0309
KNN	0.733 ± 0.0400	0.850 ± 0.0431	0.745 ± 0.0343	0.852 ± 0.0408	0.703 ± 0.0530	0.841 ± 0.0564	0.768 ± 0.0394	0.860 ± 0.0383
GNB	0.757 ± 0.0277	0.787 ± 0.0454	0.741 ± 0.0316	0.803 ± 0.0240	0.805 ± 0.0391	0.754 ± 0.1500	0.716 ± 0.0336	0.868 ± 0.1055

Para melhor ilustrar essa diferença foi apresentado, na Figura 21. Onde, é possível identificar visualmente a contribuição da inclusão da análise de imagem, reforçando a hipótese sobre a utilização DFSC(Est,Img), onde, aparentemente existe uma alteração e melhoria de cada uma das métricas.

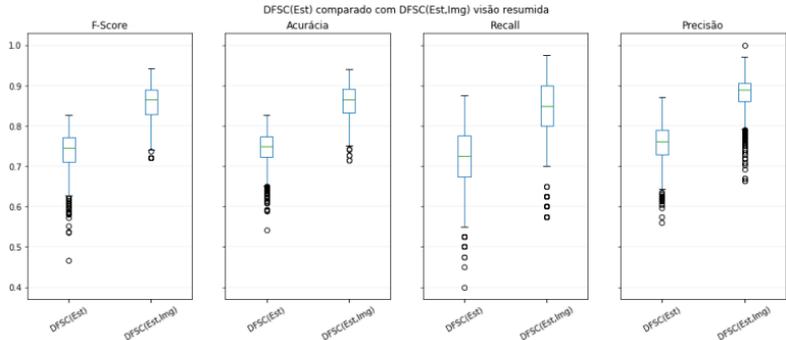


Figura 21 – Visão Geral DFSC(Est,Img)

Logo, em uma primeira análise, pode-se observar que os valores das médias de F-Score, Acurácia, Recall e Precisão obtidas com o DFSC(Est,Img) foram superiores aos valores correspondentes obtidos com o DFSC(Est). Tal fato se constitui em evidência experimental que aponta para a validade da hipótese levantada por este trabalho de que considerar imagens sobre os produtos de forma combinada com os dados estruturados de anúncios em *social commerce* pode melhorar o desempenho dos métodos de detecção de fraudes que se baseiam exclusivamente nos dados estruturados desses anúncios.

As próximas subseções apresentam uma análise detalhada ao utilizar DFSC(Est) ou DFSC(Est,Img) e sua significância estatística (obtida por meio da utilização dos conceitos descritos em 5.6) dos resultados considerando cada algoritmo de classificação e cada métrica de avaliação separadamente.

5.8.3.1 Avaliações e relevância estatística de cada uma das métricas

Para confirmar o aparente ganho é necessário que isto seja validado estatisticamente, desse modo, na Tabela 24, podem ser observadas, em resumo, para cada uma das métricas se a inclusão da imagem altera de modo significativo ou não. Nela foram apresentadas, para cada algoritmo combinador a avaliação individual de cada métrica, qual foi o teste aplicado para descobrir normalidade e seus respectivos valores. Além da conclusão se a aplicação do teste de normalidade concluiu ser uma distribuição normal ou não. A partir disso foram aplicados os Testes de Hipótese, como era condicional a cada tipo de distribuição eles são explicitados e seus respectivos valores registrados. Por fim, a última coluna registra se é possível afirmar que a inclusão da imagem exerceu influência estatisticamente significativa ou se seria necessário descartar essa hipótese podendo acontecer ao acaso. Maiores detalhes com relação a esta avaliação estatística poderá ser encontrada no apêndice C.

Desse modo, uma análise permite observar que quase todas as métricas de todos os modelos tiveram diferença significativa. Sendo assim, obteve-se evidências experimentais que reforçaram a hipótese inicial de que considerar as imagens para os modelos de detecção em *social commerce* contribuiria para a qualidade da detecção.

Tabela 24 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Combinadores Avaliados após inclusão da Imagem

Combinador	Métrica Avaliada	Teste de Normalidade (Shapiro-Wilk)	Distribuição Normal?	Teste de Hipótese	Valores no Teste de Hipótese	Estatisticamente significativa?
GAU	F-Score	(0.90, 0.000)	Não	Wilcoxon	(1647.00, 0.000)	Sim
GAU	Acurácia	(0.89, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	Recall	(0.92, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	Precisão	(0.93, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	F-Score	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	Acurácia	(0.98, 0.027)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	Recall	(0.97, 0.022)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	Precisão	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	F-Score	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	Acurácia	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	Recall	(0.97, 0.022)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	Precisão	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	F-Score	(0.98, 0.091)	Sim	T-pareado	(-49.60, 0.000)	Sim
MLP	Acurácia	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	Recall	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	Precisão	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	F-Score	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	Acurácia	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	Recall	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	Precisão	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	F-Score	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	Acurácia	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	Recall	(0.92, 0.000)	Não	Wilcoxon	(66.00, 0.000)	Sim
LDA	Precisão	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
DCT	F-Score	(0.99, 0.778)	Sim	T-pareado	(-24.12, 0.000)	Sim
DCT	Acurácia	(0.99, 0.629)	Sim	T-pareado	(-24.49, 0.000)	Sim
DCT	Recall	(0.99, 0.444)	Sim	T-pareado	(-21.18, 0.000)	Sim
DCT	Precisão	(0.99, 0.931)	Sim	T-pareado	(-19.33, 0.000)	Sim

6 CONCLUSÃO

O contínuo crescimento do volume de transações de *social commerce* tem potencializado a ocorrência de fraudes *online*. O *social commerce* é uma forma de comércio eletrônico com características de redes sociais ou ocorrida em redes sociais, onde os produtos a serem comercializados são divulgados por meio de anúncios que, em geral, contém informações estruturadas (e.g., preço, tipo de produto, marca, modelo, dentre outras) e imagens (fotos) associadas a esses produtos.

As principais soluções existentes na detecção automática de fraude em *social commerce* utilizam modelos de classificação binária (fraude x não fraude), construídos pela aplicação de algoritmos de aprendizado de máquina, e podem ser divididas em dois grupos, em função do tipo de informação extraída dos anúncios. Enquanto no primeiro grupo, estão os modelos que utilizam apenas informações estruturadas, no segundo estão os que se limitam ao uso de imagens. Independente do grupo a que pertencem, todas essas soluções se restringem a considerar apenas parte das informações que podem conter indícios de presença de fraude nos anúncios. Por exemplo, as soluções do primeiro grupo ignoram potenciais indicativos de fraude nas imagens, que complementam as informações sobre o objeto comercializado.

Diante do exposto, o presente trabalho levantou a hipótese de que combinar informações estruturadas dos anúncios com informações extraídas das imagens associadas a esses anúncios, poderia elevar a acurácia dos modelos de detecção de fraudes.

A fim de comprovar esta hipótese, o trabalho propôs um método que extrai informações das imagens, por meio de *Deep Learning*, e as combina com informações estruturadas associadas aos anúncios para gerar os modelos de detecção de fraude.

Para as combinações foi proposto um método que permite quatro alternativas de utilização independentes: DFSC(Est), DFSC(Img), DFSC(Est+Img), DFSC(Est,Img). As duas últimas atuando na combinação de atributos de fato. Se, por um lado, o DFSC(Est+Img) atua em dados estruturados após o enriquecimento dos dados da imagem, por outro, DFSC(Est,Img) permite a união após o processamento individual feito por modelos independentes, viabilizado por um algoritmo classificador.

Os experimentos realizados em uma base de anúncios real de uma grande empresa de *Social Commerce* apresentaram ganhos com a aplicação dos métodos em F-Score, com DFSC(Est+Img) ganhos de até 7% e com DFSC(Est,Img) ganhos de até 20%, ambos com diferenças estatísticas significativas. Desse modo, em ambas as abordagens foi possível observar indicativos de ganhos de robustez com os modelos gerados pelo método proposto, obtendo, desta forma, evidências experimentais que apontam para a validade da hipótese

levantada. Infelizmente, devido a termo de sigilo assinado, maiores informações sobre a empresa e código desenvolvido não podem ser divulgados.

Assim sendo, as principais contribuições deste trabalho foram o método proposto, o detalhamento dos experimentos realizados e os resultados promissores obtidos pelo método nesses experimentos.

Entre as principais limitações desta dissertação e que deverão ser objeto de iniciativas de trabalhos futuros estão: reduzir a dimensão do problema, selecionando os atributos mais relevantes para a tarefa de detecção; testar outras formas de combinar *modelos Est* e *modelos Img*, por exemplo, por meio da utilização de múltiplos modelos combinadores no lugar de apenas um; e, em DFSC(Est,Img) incluir o SVM (Máquinas de Vetores de Suporte) para classificar os dados estruturados, uma vez que para este experimento, esse algoritmo não foi considerado por não atender o requisito de gerar uma classificação probabilística na sua saída. Explorar outros atributos não estruturados como comentários do usuário ou texto da descrição do anúncio, bem como verificar a extração de potenciais marcas d'água ou textos nas imagens e considerar elas em separado agregaria para detecção de fraude.

REFERÊNCIAS

- 1 FREEH, L. *Citação do Louis Freeh*. 2020. <<https://www.fraud-magazine.com/article.aspx?id=4294981971>>.
- 2 FBI. 2019 Annual Internet Crime Report. *Federal Bureau of Investigation - Internet Crime Complaint Center*, v. 0, 2019. Disponível em: <https://pdf.ic3.gov/2019{_}IC3Report.>
- 3 BEALS, M.; DELIEMA, M.; DEEVY, M. Framework for a Taxonomy of Fraud. *Washington DC: Stanford Longevity Center/FINRA Financial Investor Education Foundation/Fraud Research Center*, v. 25, n. July, p. 40, 2015. Disponível em: <<http://longevity3.stanford.edu/wp-content/uploads/2015/11/Full-Taxonomy-report.pdf>>.
- 4 CARROS só pra rodar: entenda o que significa e como funciona. *Notícias Automotivas*, Notícias Automotivas, v. 0, 2018. Disponível em: <<https://www.noticiasautomotivas.com.br/carros-so-pra-rodar>>.
- 5 SEBRAE. *O que você precisa saber sobre comércio eletrônico*. 2017. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/artigos/o-que-voce-precisa-saber-sobre-comercio-eletronico,1a19d1eb00ad2410VgnVCM100000b272010aRCRD>>.
- 6 MAVLANOVA, T.; BENBUNAN-FICH, R.; LANG, G. The role of external and internal signals in E-commerce. *Decision Support Systems*, Elsevier B.V., v. 87, p. 59–68, 2016. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2016.04.009>>.
- 7 DOBREVA, K. *Global E-commerce Trends and Statistics*. <https://amasty.com>, 2018. 26/04/2018. Disponível em: <<https://amasty.com/blog/wp-content/uploads/2018/02/GLOBAL-E-COMMERCE-TRENDS-AND-STATISTICS.pdf>>.
- 8 LARGHI, N. *Compras online crescem na quarentena veja produtos mais vendidos*. 2020. Disponível em: <<https://valorinveste.globo.com/objetivo/gastar-bem/noticia/2020/05/15/compras-on-line-crescem-na-quarentena-veja-produtos-mais-vendidos.ghtml>>.
- 9 STATISTA, D. Number of monthly active facebook users worldwide as of 3rd quarter 2020. *Statista*, Statista, v. 0, 2020. Disponível em: <<https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>>.
- 10 STATISTA, D. Leading countries based on instagram audience size as of october 2020. *Statista*, Statista, v. 0, 2020. Disponível em: <<https://www.statista.com/statistics/578364/countries-with-most-instagram-users>>.
- 11 STATISTA, D. Monthly active facebook users worldwide. *Statista*, Statista, v. 0, 2018. Disponível em: <<https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>>.
- 12 CURTY, R. G.; ZHANG, P. Social commerce: Looking back and forward. *Proceedings of the ASIST Annual Meeting*, v. 48, p. 1–10, 2011. ISSN 15508390.

- 13 ZHANG, K. Z. K.; BENYOUCEF, M. Consumer behavior in social commerce : A literature review. *Decision Support Systems*, Elsevier B.V., v. 86, p. 95–108, 2016. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2016.04.001>>.
- 14 STATISTA, D. Leading online shopping channels in brazil in 2019. *Statista*, Statista, v. 0, 2020. Disponível em: <<https://www.statista.com/statistics/444015/preferred-digital-retailer-access-brazil>>.
- 15 ABDALLAH, A.; MAAROF, M. A.; ZAINAL, A. Fraud detection system: A survey. *Journal of Network and Computer Applications*, v. 68, p. 90–113, 2016. ISSN 10958592.
- 16 FBI. 2017 Annual Internet Crime Report. *Federal Bureau of Investigation - Internet Crime Complaint Center*, v. 0, 2017. Disponível em: <https://pdf.ic3.gov/2017/{_}IC3Report.>
- 17 FEARN-BANKS, K. *Crisis communications: A casebook approach*. [S.l.: s.n.], 2016. 438 p.
- 18 WHITTY, M. T.; BUCHANAN, T. The online romance scam: A serious cybercrime. *CyberPsychology, Behavior, and Social Networking*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 15, n. 3, p. 181–183, 2012.
- 19 SAHIN, Y.; BULKAN, S.; DUMAN, E. Expert Systems with Applications A cost-sensitive decision tree approach for fraud detection. *Expert Systems With Applications*, Elsevier Ltd, v. 40, n. 15, p. 5916–5923, 2013. ISSN 0957-4174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2013.05.021>>.
- 20 CHEUNG, M.; SHE, J.; LIU, L. Deep Learning-based Online Counterfeit-seller Detection. *INFOCOM - IEEE International Conference on Computer Communications*, IEEE, v. 1, p. 51–56, 2018.
- 21 CARTA, S.; FENU, G.; Reforgiato Recupero, D.; SAIA, R. Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. *Journal of Information Security and Applications*, Elsevier Ltd, v. 46, n. 1, p. 13–22, 2019. ISSN 22142126.
- 22 ALMENDRA, V. Finding the needle: A risk-based ranking of product listings at online auction sites for non-delivery fraud prediction. *Expert Systems with Applications*, Elsevier Ltd, v. 40, n. 12, p. 4805–4811, 2013. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2013.02.027>>.
- 23 FORD, B. J.; XU, H.; VALOVA, I. A real-time self-adaptive classifier for identifying suspicious bidders in online auctions. *Computer Journal*, v. 56, n. 5, p. 646–663, 2013. ISSN 00104620.
- 24 ZHOU, L.; CHEN, W. B.; ZHANG, C. Authorship detection and encoding for eBay images. *Multimedia Data Engineering Applications and Processing*, v. 1, p. 20–34, 2013.
- 25 YANG, L.; CHEN, W. B.; ZHANG, C.; JOHNSTONE, J. K.; GAO, S.; WARNER, G. Profiling online auction sellers using image-editing styles. *IEEE Multimedia*, v. 19, n. 1, p. 29–39, 2012. ISSN 1070986X.

- 26 CHEUNG, M.; SHE, J.; WANG, N. Characterizing User Connections in Social Media through User Shared Image. *IEEE Transactions on Big Data*, v. 0, n. 0, p. 1–1, 2017. ISSN 2332-7790. Disponível em: <<http://ieeexplore.ieee.org/document/8067486/>>.
- 27 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015. ISSN 14764687.
- 28 TORREY, L.; SHAVLIK, J. Transfer learning. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. [S.l.]: IGI global, 2010. v. 1, p. 242–264.
- 29 SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 4510–4520.
- 30 KUNCHEVA, L. I. *Base Classifiers*. [S.l.: s.n.], 2004. 45–94 p. ISSN 1045-9227. ISBN 9786468600.
- 31 KÜTZ, M. *Introduction to E-Commerce: Combining Business and Information Technology*. 1st ed. ed. [S.l.]: bookboon, 2016. 209 p. ISBN 9788740315202.
- 32 TURBAN, E.; KING, D. *Electronic Commerce: Managerial and Social Networks Perspectives*. [S.l.: s.n.], 2012. 68 p. ISBN 9783319100906.
- 33 TURBAN, E.; KING, D.; LEE, J. K.; LIANG, T.-P.; TURBAN, D. C. *Electronic Commerce*. 8th. ed. Cham: Springer International Publishing, 2015. 68 p. (Springer Texts in Business and Economics). ISBN 978-3-319-10090-6. Disponível em: <<http://link.springer.com/10.1007/978-3-319-10091-3>>.
- 34 WANG, C. N. S. Y.-S. U.; ZHANG, P. S. U. The Evolution of Social Commerce: The People , Business , Technology , and Information Dimensions and Information Dimensions. *Communications of the Association for Information Systems*, v. 31, n. August, p. 5, 2012.
- 35 GOLDSCHMIDT, R.; PASSOS, E. *Data mining: um guia prático*. [S.l.]: Gulf Professional Publishing, 2005.
- 36 GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.
- 37 PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.
- 38 VLASSELAER, V. V.; BRAVO, C.; CAELEN, O.; ELIASSI-RAD, T.; AKOGLU, L.; SNOECK, M.; BAESENS, B. APATE : A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, Elsevier B.V., v. 75, p. 38–48, 2015. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2015.04.013>>.
- 39 JHAVERI, M. H.; CETIN, O.; GAÑÁN, C.; MOORE, T.; EETEN, M. V. Abuse Reporting and the Fight Against Cybercrime. *ACM Computing Surveys*, v. 49, n. 4, p. 1–27, 2017. ISSN 03600300. Disponível em: <<http://dl.acm.org/citation.cfm?doid=3022634.3003147>>.

- 40 NGAI, E. W.; HU, Y.; WONG, Y. H.; CHEN, Y.; SUN, X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, Elsevier B.V., v. 50, n. 3, p. 559–569, 2011. ISSN 01679236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2010.08.006>>.
- 41 LAI, F.; LI, D.; HSIEH, C.-t. Fighting identity theft : The coping perspective. *Decision Support Systems*, Elsevier B.V., v. 52, n. 2, p. 353–363, 2012. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2011.09.002>>.
- 42 RATHORE, S.; KUMAR, P.; LOIA, V.; JEONG, Y.-s.; HYUK, J. Social network security : Issues , challenges , threats , and solutions. *Information Sciences*, Elsevier Inc., v. 421, p. 43–69, 2017. ISSN 0020-0255. Disponível em: <<http://dx.doi.org/10.1016/j.ins.2017.08.063>>.
- 43 ZHAO, J.; LAU, R. Y. K.; ZHANG, W.; ZHANG, K.; CHEN, X.; TANG, D. Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce. *Decision Support Systems*, Elsevier B.V., v. 86, p. 109–121, 2016. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2016.04.003>>.
- 44 TSANG, S.; KOH, Y. S.; DOBBIE, G.; ALAM, S. SPAN: Finding collaborative frauds in online auctions. *Knowledge-Based Systems*, Elsevier B.V., v. 71, p. 389–408, 2014. ISSN 09507051. Disponível em: <<http://dx.doi.org/10.1016/j.knosys.2014.08.016>>.
- 45 CARNEIRO, N.; FIGUEIRA, G.; COSTA, M. A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, Elsevier B.V., v. 95, p. 91–101, 2017. ISSN 0167-9236. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2017.01.002>>.
- 46 WIMMER, H.; YOON, V. Y. Counterfeit product detection: Bridging the gap between design science and behavioral science in information systems research. *Decision Support Systems*, Elsevier B.V., v. 104, p. 1–12, 2017. ISSN 0167-9236. Disponível em: <<https://doi.org/10.1016/j.dss.2017.09.005>>.
- 47 ALARIFI, A.; ALSALEH, M.; AL-SALMAN, A. Twitter turing test : Identifying social machines R. *Information Sciences*, Elsevier Inc., v. 372, p. 332–346, 2016. ISSN 0020-0255. Disponível em: <<http://dx.doi.org/10.1016/j.ins.2016.08.036>>.
- 48 YASAMI, Y.; SAFAEI, F. A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks. *Computer Communications*, Elsevier B.V., v. 100, p. 52–64, 2017. ISSN 0140-3664. Disponível em: <<http://dx.doi.org/10.1016/j.comcom.2016.11.010>>.
- 49 WU, S.-h.; CHOU, M.-j.; TSENG, C.-h.; LEE, Y.-j.; CHEN, K.-t. Detecting \textit{In Situ} Identity Fraud on Social Network Services: A Case Study With Facebook. *IEEE Systems Journal*, v. 11, n. 4, p. 1–12, 2015. ISSN 1932-8184. Disponível em: <<http://ieeexplore.ieee.org/document/7362137/>>.
- 50 WU, S.-H.; CHOU, M.-J.; TSENG, C.-H.; LEE, Y.-J.; CHEN, K.-T. Detecting In Situ Identity Fraud on Social Network Services: A Case Study With Facebook. *IEEE Systems Journal*, v. 11, n. 4, p. 2432–2443, 2017. Disponível em: <http://www.ieee.org/publications{_}standards/publications/rights/index.h>.

- 51 ADEWOLE, K. S.; ANUAR, N. B.; KAMSIN, A.; VARATHAN, K. D.; ABDUL, S. Malicious accounts : Dark of the social networks. *Journal of Network and Computer Applications*, v. 79, n. September 2016, p. 41–67, 2017.
- 52 AKHTER, M. I.; AHAMAD, M. G. Detecting Telecommunication Fraud using Neural Networks through Data Mining. *International Journal of Scientific & Engineering Research*, v. 3, n. 3, p. 1–5, 2012. ISSN 2229-5518. Disponível em: <<http://www.ijser.org>>.
- 53 BRENNAN, P.; HOFFMAN, M. *A comprehensive survey of methods for overcoming the class imbalance problem in fraud detection*. Dublin, Ireland: [s.n.], 2012. 1–107 p. (Computing, June).
- 54 PHUA, C.; SMITH-MILES, K.; LEE, V.; GAYLER, R. Resilient identity crime detection. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 24, n. 3, p. 533–546, 2012. ISSN 10414347.
- 55 Dal Pozzolo, A.; CAELEN, O.; WATERSCHOOT, S.; BONTEMPI, G. Racing for unbalanced methods selection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 8206 LNCS, n. 1, p. 24–31, 2013. ISSN 03029743.
- 56 CARCILLO, F.; Dal Pozzolo, A.; Le Borgne, Y. A.; CAELEN, O.; MAZZER, Y.; BONTEMPI, G. SCARFF: A scalable framework for streaming credit card fraud detection with spark. *Information Fusion*, Elsevier B.V., v. 41, p. 182–194, 2018. ISSN 15662535.
- 57 AUGUSTIN, S.; GAISSER, C.; KNAUER, J.; MASSOTH, M.; PIEJKO, K.; RIHM, D.; WIENS, T. Telephony fraud detection in next generation networks. *CERC2012*, v. 0, p. 13–20, 2012.
- 58 CHANG, W. H.; CHANG, J. S. An effective early fraud detection method for online auctions. *Electronic Commerce Research and Applications*, Elsevier B.V., v. 11, n. 4, p. 346–360, 2012. ISSN 15674223. Disponível em: <<http://dx.doi.org/10.1016/j.elerap.2012.02.005>>.
- 59 JHA, S.; GUILLEN, M.; Christopher Westland, J. Employing transaction aggregation strategy to detect credit card fraud. *Expert Systems with Applications*, Elsevier Ltd, v. 39, n. 16, p. 12650–12657, 2012. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2012.05.018>>.
- 60 LEI, L. Card fraud detection by inductive learning and evolutionary algorithm. *Proceedings - 2012 6th International Conference on Genetic and Evolutionary Computing, ICGEC 2012*, IEEE, v. 0, n. 156, p. 384–388, 2012. Disponível em: <<https://ieeexplore.ieee.org/document/6457283/>>.
- 61 Han Tao; Liu Zhixin; Song Xiaodong. Insurance fraud identification research based on fuzzy support vector machine with dual membership. *2012 International Conference on Information Management, Innovation Management and Industrial Engineering*, IEEE, v. 3, n. 0, p. 457–460, 2012.
- 62 TSANG, S.; DOBBIE, G.; KOH, Y. S. Evaluating fraud detection algorithms using an auction data generator. *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, IEEE, v. 1, n. 1, p. 332–339, 2012.

- 63 WONG, N.; RAY, P.; STEPHENS, G.; LEWIS, L. Artificial immune systems for the detection of credit card fraud: An architecture, prototype and preliminary results. *Information Systems Journal*, v. 22, n. 1, p. 53–76, 2012. ISSN 13501917.
- 64 AL-FAYOUMI, M. A.; SHILBAYEH, N. F. Cloning SIM cards usability reduction in mobile networks. *Journal of Network and Systems Management*, v. 22, n. 2, p. 259–279, 2014. ISSN 10647570.
- 65 CHEN, S.; GANGOPADHYAY, A. A novel approach to uncover health care frauds through spectral analysis. *Proceedings - 2013 IEEE International Conference on Healthcare Informatics, ICHI 2013*, IEEE, v. 1, n. 1, p. 499–504, 2013.
- 66 LIU, Q.; STATES, U.; VASARHELYI, M.; STATES, U. Healthcare fraud detection : A survey and a clustering model incorporating Geo-location information. *29th World Continuous Auditing and Reporting Symposium (29wcars)*, v. 1, n. 1, p. 10, 2013. Disponível em: <<http://raw.rutgers.edu/docs/wcars/29wcars/HealthcarefrauddetectionAsurveyandaclusteringmodelincorporatingGeo-locationinformation.pdf>>.
- 67 MALEKIAN, D.; HASHEMI, M. R. An adaptive profile based fraud detection framework for handling concept drift. *2013 10th International ISC Conference on Information Security and Cryptology, ISCISC 2013*, IEEE, v. 1, n. 1, p. 1–6, 2013.
- 68 SAHIN, Y.; BULKAN, S.; DUMAN, E. A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, Elsevier Ltd, v. 40, n. 15, p. 5916–5923, 2013. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2013.05.021>>.
- 69 CHANG, J. S.; CHANG, W. H. Analysis of fraudulent behavior strategies in online auctions for detecting latent fraudsters. *Electronic Commerce Research and Applications*, Elsevier B.V., v. 13, n. 2, p. 79–97, 2014. ISSN 15674223. Disponível em: <<http://dx.doi.org/10.1016/j.elerap.2013.10.004>>.
- 70 Dal Pozzolo, A.; CAELEN, O.; Le Borgne, Y. A.; WATERSCHOOT, S.; BONTEMPI, G. Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, Elsevier Ltd, v. 41, n. 10, p. 4915–4928, 2014. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2014.02.026>>.
- 71 SARAVANAN, P.; SUBRAMANIASWAMY, V.; SIVARAMAKRISHNAN, N.; PRAKASH, M. A.; ARUNKUMAR, T. Data mining approach for subscription-fraud detection in telecommunication sector. *Contemporary Engineering Sciences*, v. 7, n. 11, p. 515–522, 2014.
- 72 SEEJA, K. R.; ZAREAPOOR, M. FraudMiner: A novel credit card fraud detection model based on frequent itemset mining. *Scientific World Journal*, v. 2014, n. 1, p. 10, 2014. ISSN 1537744X.
- 73 TSAI, Y. H.; KO, C. H.; LIN, K. C. Using commonKADS method to build prototype system in medical insurance fraud detection. *Journal of Networks*, v. 9, n. 7, p. 1798–1802, 2014. ISSN 17962056.
- 74 TSANG, S.; KOH, Y. S.; DOBBIE, G.; ALAM, S. Detecting online auction shilling frauds using supervised learning. *Expert Systems with Applications*, Elsevier Ltd, v. 41,

- n. 6, p. 3027–3040, 2014. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2013.10.033>>.
- 75 ZANIN, M.; ROMANCE, M.; MORAL, S.; CRIADO, R. Credit Card Fraud Detection through Parenchitic Network Analysis. *Complexity*, v. 2018, n. 1, p. 9, 2018. ISSN 10990526.
- 76 COUNTERFEIT merchandise. *Canadian Anti-Fraud Centre*, Canadian Anti-Fraud Centre, v. 0, 2020. Disponível em: <<https://www.antifraudcentre-centreantifraude.ca/scams-fraudes/merchandise-marchandises-eng.htm>>.
- 77 "CARROS só para rodar" - Conheça a feira do rolo virtual que cresce a cada dia. *motor1Uol*, UOL, v. 0, 2017. Disponível em: <<https://motor1.uol.com.br/news/140466/carros-so-para-rodar-conheca-a-feira-do-rolo-virtual-que-cresce-a-cada-dia/>>.
- 78 ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Taylor & Francis Group, v. 46, n. 3, p. 175–185, 1992.
- 79 FREEDMAN, D. A. *Statistical models: theory and practice*. [S.l.]: cambridge university press, 2009.
- 80 HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007.
- 81 BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- 82 QUINLAN, J. R. Simplifying decision trees. *International journal of man-machine studies*, Elsevier, v. 27, n. 3, p. 221–234, 1987.
- 83 MCLACHLAN, G. J. *Discriminant analysis and statistical pattern recognition*. [S.l.]: John Wiley & Sons, 2004. v. 544.
- 84 JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. *arXiv preprint arXiv:1302.4964*, v. 1, p. 8, 2013. Disponível em: <<https://arxiv.org/abs/1302.4964v1>>.
- 85 CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995.
- 86 HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- 87 DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *IEEE. 2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255.
- 88 HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

APÊNDICE A – BIBLIOTECAS USADAS

Python (Python 3.6.10 :: Anaconda, Inc)

alabaster==0.7.12
anaconda-client==1.7.2
anaconda-project==0.8.3
argh==0.26.2
asn1crypto==1.3.0
astroid==2.4.2
astropy==4.0
atomicwrites==1.3.0
attrs==19.3.0
Automat==20.2.0
autopep8==1.4.4
autovizwidget==0.16.0
awscli==1.18.169
Babel==2.8.0
backcall==0.1.0
backports.shutil-get-terminal-size==1.0.0
bcrypt==3.2.0
beautifulsoup4==4.8.2
bitarray==1.2.1
bkcharts==0.2
bleach==3.2.1
bokeh==1.4.0
boto==2.49.0
boto3==1.16.9
botocore==1.19.9
Bottleneck==1.3.2
cached-property==1.5.2
certifi==2020.6.20
cffi==1.14.0
chardet==3.0.4
Click==7.0
cloudpickle==1.3.0
clyent==1.2.2
colorama==0.4.3

contextlib2==0.6.0.post1
cryptography==2.8
cyclerc==0.10.0
Cython==0.29.15
cytoolz==0.10.1
dask==2.11.0
decorator==4.4.1
defusedxml==0.6.0
diff-match-patch==20181111
distributed==2.11.0
distro==1.5.0
docker==4.3.1
docker-compose==1.27.4
dockerpty==0.4.1
docopt==0.6.2
docutils==0.15.2
entrypoints==0.3
environment-kernels==1.1.1
et-xmlfile==1.0.1
fastcache==1.1.0
filelock==3.0.12
flake8==3.7.9
Flask==1.1.1
fsspec==0.6.2
future==0.18.2
gevent==1.4.0
glob2==0.7
gmpy2==2.0.8
google-pasta==0.2.0
greenlet==0.4.15
h5py==2.10.0
hdijupyterutils==0.16.0
HeapDict==1.0.1
html5lib==1.0.1
hypothesis==5.5.4
idna==2.8
imageio==2.6.1
imagesize==1.2.0
importlib-metadata==1.5.0

intervaltree==3.0.2
ipykernel==5.1.4
ipyparallel==6.3.0
ipython==7.12.0
ipython-genutils==0.2.0
ipywidgets==7.5.1
isort==4.3.21
itsdangerous==1.1.0
jdcal==1.4.1
jedi==0.14.1
jeepney==0.4.2
Jinja2==2.11.1
jmespath==0.10.0
joblib==0.14.1
json5==0.9.1
jsonschema==3.2.0
jupyter==1.0.0
jupyter-client==5.3.4
jupyter-console==6.1.0
jupyter-core==4.6.1
jupyterlab==1.2.6
jupyterlab-server==1.0.6
keyring==21.1.0
kiwisolver==1.1.0
lazy-object-proxy==1.4.3
libarchive-c==2.8
lief==0.9.0
llvmlite==0.31.0
locket==0.2.0
lxml==4.6.1
MarkupSafe==1.1.1
matplotlib==3.1.3
mccabe==0.6.1
mistune==0.8.4
mkl-fft==1.0.15
mkl-random==1.1.0
mkl-service==2.3.0
mock==4.0.1
more-itertools==8.2.0

mpmath==1.1.0
msgpack==0.6.1
multipledispatch==0.6.0
nb-conda==2.2.1
nb-conda-kernels==2.2.4
nbconvert==5.6.1
nbformat==5.0.4
networkx==2.4
nltk==3.4.5
nose==1.3.7
notebook==6.0.3
numba==0.48.0
numexpr==2.7.1
numpy==1.18.1
numpydoc==0.9.2
nvidia-ml-py==10.418.84
olefile==0.46
opencv-python==4.2.0.32
openpyxl==3.0.3
packaging==20.1
pandas==1.0.1
pandocfilters==1.4.2
paramiko==2.7.2
parso==0.5.2
partd==1.1.0
path==13.1.0
pathlib2==2.3.5
pathtools==0.1.2
patsy==0.5.1
pep8==1.7.1
pexpect==4.8.0
pickleshare==0.7.5
Pillow==7.2.0
pkginfo==1.5.0.1
plotly==4.12.0
pluggy==0.13.1
ply==3.11
prometheus-client==0.7.1
prompt-toolkit==3.0.3

protobuf==3.11.4
protobuf3-to-dict==0.1.5
psutil==5.6.7
psycopg2==2.7.5
PTable==0.9.2
ptyprocess==0.6.0
py==1.8.1
py4j==0.10.7
pyasn1==0.4.8
pycodestyle==2.5.0
pycosat==0.6.3
pyparser==2.19
pycrypto==2.6.1
pycurl==7.43.0.5
pydocstyle==4.0.1
pyflakes==2.1.1
pygal==2.4.0
Pygments==2.5.2
pykerberos==1.2.1
pylint==2.5.3
PyNaCl==1.4.0
pyodbc===4.0.0-unsupported
pyOpenSSL==19.1.0
pyparsing==2.4.6
pyrsistent==0.15.7
PySocks==1.7.1
pyspark==2.3.4
pytest==5.3.5
pytest-arraydiff==0.3
pytest-astropy==0.8.0
pytest-astropy-header==0.1.2
pytest-doctestplus==0.5.0
pytest-openfiles==0.4.0
pytest-remotedata==0.3.2
python-dateutil==2.8.1
python-dotenv==0.15.0
python-jsonrpc-server==0.3.4
python-language-server==0.31.7
pytz==2019.3

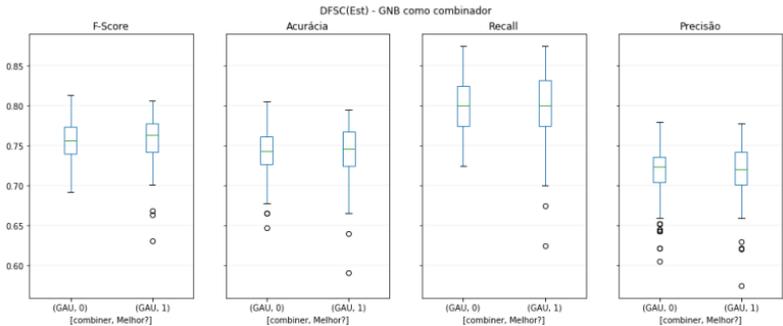
PyWavelets==1.1.1
pyxdg==0.26
PyYAML==5.3.1
pyzmq==18.1.1
QDarkStyle==2.8
QtAwesome==0.6.1
qtconsole==4.6.0
QtPy==1.9.0
requests==2.22.0
requests-kerberos==0.12.0
retrying==1.3.3
rope==0.16.0
rsa==4.5
Rtree==0.9.3
ruamel-yaml==0.15.87
s3fs==0.4.2
s3transfer==0.3.3
sagemaker==2.16.1
sagemaker-pyspark==1.4.1
scikit-image==0.16.2
scikit-learn==0.22.1
scipy==1.4.1
seaborn==0.10.0
SecretStorage==3.1.2
Send2Trash==1.5.0
simplegeneric==0.8.1
singledispatch==3.4.0.3
six==1.14.0
smdebug-rulesconfig==0.1.5
snowballstemmer==2.0.0
sortedcollections==1.1.2
sortedcontainers==2.1.0
soupsieve==1.9.5
sparkmagic==0.15.0
Sphinx==3.0.4
sphinxcontrib-applehelp==1.0.1
sphinxcontrib-devhelp==1.0.1
sphinxcontrib-htmlhelp==1.0.2
sphinxcontrib-jsmath==1.0.1

sphinxcontrib-qthelp==1.0.2
sphinxcontrib-serializinghtml==1.1.3
sphinxcontrib-websupport==1.2.0
spyder==4.0.1
spyder-kernels==1.8.1
SQLAlchemy==1.3.13
statsmodels==0.11.0
sympy==1.5.1
tables==3.6.1
tblib==1.6.0
terminado==0.8.3
testpath==0.4.4
texttable==1.6.3
toml==0.10.1
toolz==0.10.0
tornado==6.0.3
tqdm==4.42.1
traitlets==4.3.3
typed-ast==1.4.1
ujson==1.35
unicodescv==0.14.1
urllib3==1.25.10
watchdog==0.10.2
wcwidth==0.1.8
webencodings==0.5.1
websocket-client==0.57.0
Werkzeug==1.0.0
widgetsnextension==3.5.1
wrapt==1.11.2
wurlitzer==2.0.0
xlrd==1.2.0
XlsxWriter==1.2.7
xlwt==1.3.0
yapf==0.28.0
zict==1.0.0
zipp==2.2.0

APÊNDICE B – DETALHAMENTO RELEVÂNCIA ESTATÍSTICA DFSC(EST+IMG) PARA CADA UMA DAS MÉTRICAS

Utilizando como combinador o algoritmo GNB, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 22 que:

Figura 22 – DFSC(Est) - Usando GNB como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1499, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.983 com p-value: 0.1499 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.941 com p-value: 0.0000

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0003, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.952 com p-value: 0.0003 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.931 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

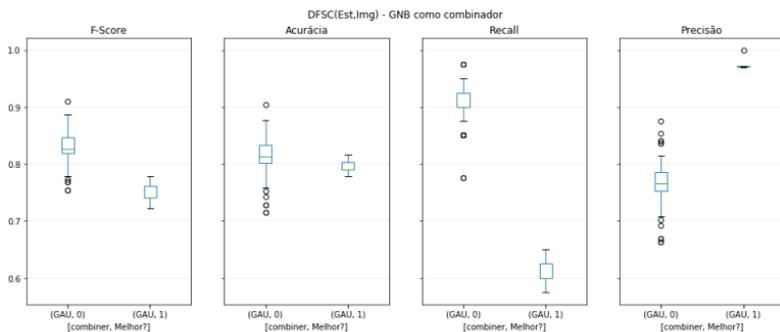
Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se:

Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0001, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.944 com p-value: 0.0001 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.895 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.921 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.930 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 23 que:

Figura 23 – DFSC(Est,Img) - Usando GNB como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0005, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.954 com p-value: 0.0005 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.833 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando

o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.939 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.832 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.870 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.829 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0007, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.956 com p-value: 0.0007 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.255 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Utilizando como combinador o algoritmo KNN, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 24 que:

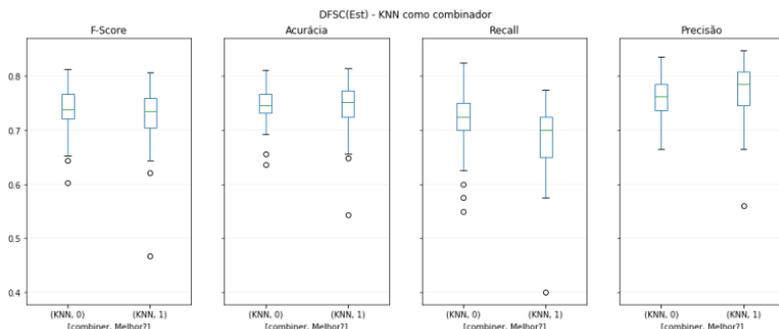


Figura 24 – DFSC(Est) - Usando KNN como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor

Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0039, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.966 com p-value: 0.0039 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.872 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0291, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.976 com p-value: 0.0291 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.906 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

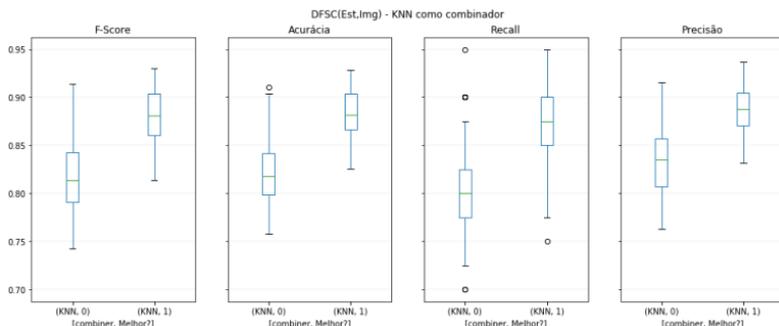
Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.939 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como

uma Distribuição Normal depois 0.891 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1510, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.984 com p-value: 0.1510 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.929 com p-value: 0.0000

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 25 que:

Figura 25 – DFSC(Est,Img) - Usando KNN como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.2471, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.986 com p-value: 0.2471 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0026, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.964 com p-value: 0.0026

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1335, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.983 com p-value: 0.1335 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0049, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.967 com p-value: 0.0049

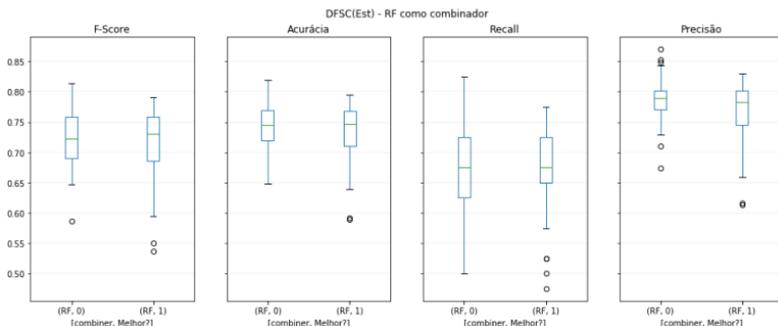
Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se:

Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0052, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.967 com p-value: 0.0052 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.933 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.4130, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.989 com p-value: 0.4130 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0551, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.979 com p-value: 0.0551

Utilizando como combinador o algoritmo RF, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 26 que:

Figura 26 – DFSC(Est) - Usando RF como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0428, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.978 com p-value: 0.0428 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.926 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

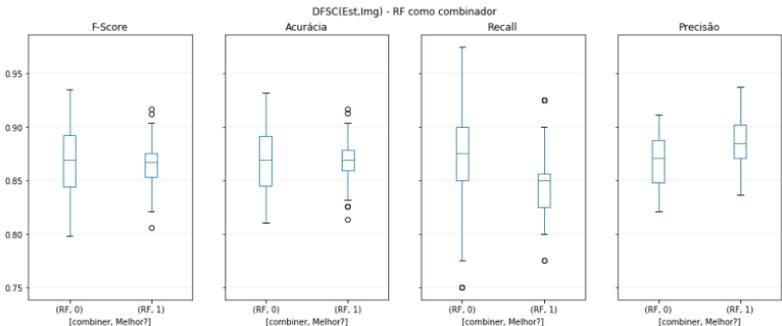
Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.3022, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.987 com p-value: 0.3022 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.925 com p-value: 0.0000

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0030, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.965 com p-value: 0.0030 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0001, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.944 com p-value: 0.0001 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0517, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.979 com p-value: 0.0517 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.902 com p-value: 0.0000

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 27 que:

Figura 27 – DFSC(Est,Img) - Usando RF como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o

Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.3164, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.987 com p-value: 0.3164 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0699, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.980 com p-value: 0.0699

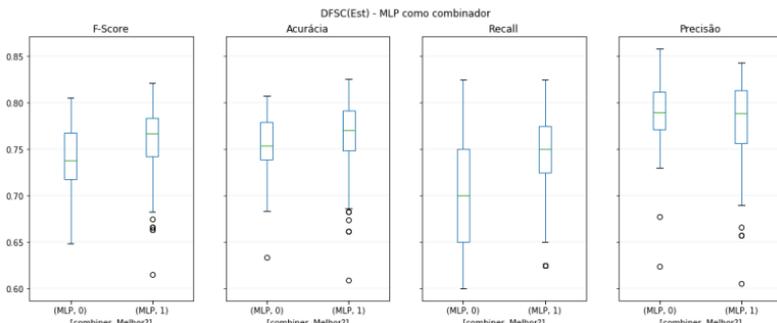
Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1304, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.983 com p-value: 0.1304 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1483, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.983 com p-value: 0.1483

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0045, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.967 com p-value: 0.0045 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.933 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0001, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.945 com p-value: 0.0001 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0465, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.978 com p-value: 0.0465 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Utilizando como combinador o algoritmo MLP, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 28 que:

Figura 28 – DFSC(Est) - Usando MLP como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.1245, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.983 com p-value: 0.1245 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.929 com p-value: 0.0000

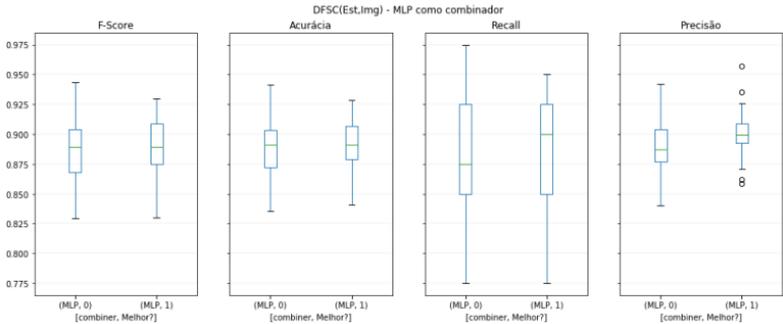
Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0024, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.963 com p-value: 0.0024 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.917 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0009, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.958 com p-value: 0.0009 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.938 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.921 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.914 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 29 que:

Figura 29 – DFSC(Est,Img) - Usando MLP como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.4094, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.988 com p-value: 0.4094 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0093, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.970 com p-value: 0.0093

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.5358, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.990 com p-value: 0.5358 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0333, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.976 com p-value: 0.0333

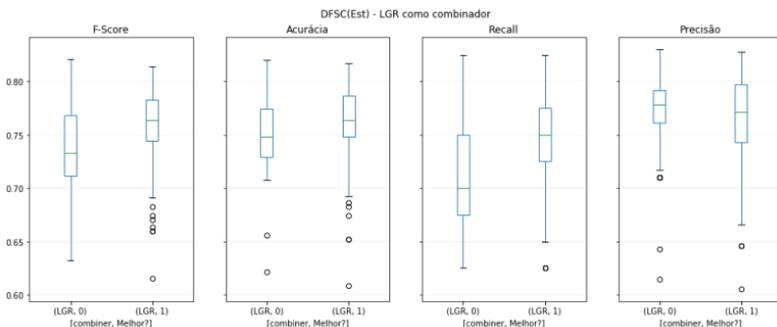
Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0016, logo podemos dizer que não

se comporta como uma Distribuição Normal antes 0.961 com p-value: 0.0016 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.926 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.5995, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.991 com p-value: 0.5995 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0147, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.973 com p-value: 0.0147

Utilizando como combinador o algoritmo LGR, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 30 que:

Figura 30 – DFSC(Est) - Usando LGR como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0001, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.944 com p-value: 0.0001 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.907 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se:

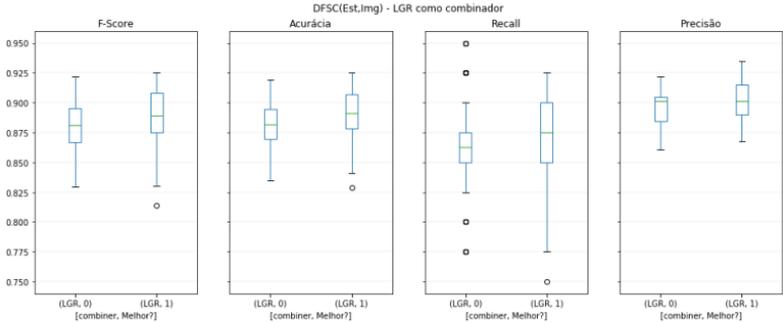
Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.935 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.900 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.901 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.900 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.881 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.918 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 31 que:

Figura 31 – DFSC(Est,Img) - Usando LGR como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0077, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.969 com p-value: 0.0077 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0014, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.961 com p-value: 0.0014 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0071, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.969 com p-value: 0.0071 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0061, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.968 com p-value: 0.0061 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0002, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.948 com p-value: 0.0002 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.898 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando

o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0010, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.959 com p-value: 0.0010 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.936 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Utilizando como combinador o algoritmo LDA, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura B que:

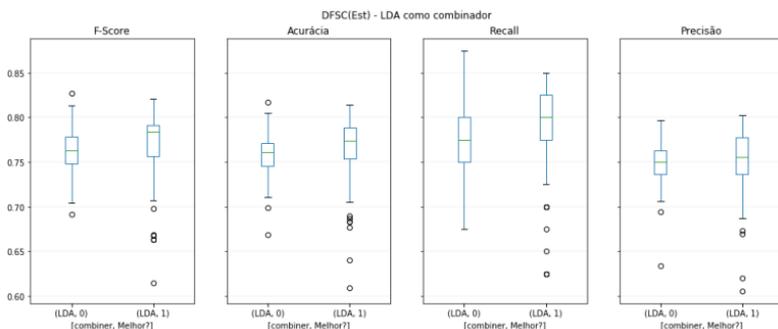


Figura 32 – DFSC(Est) - Usando LDA como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor

Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.4756, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.989 com p-value: 0.4756 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.847 com p-value: 0.0000

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0182, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.974 com p-value: 0.0182 usando o Teste de Shapiro-

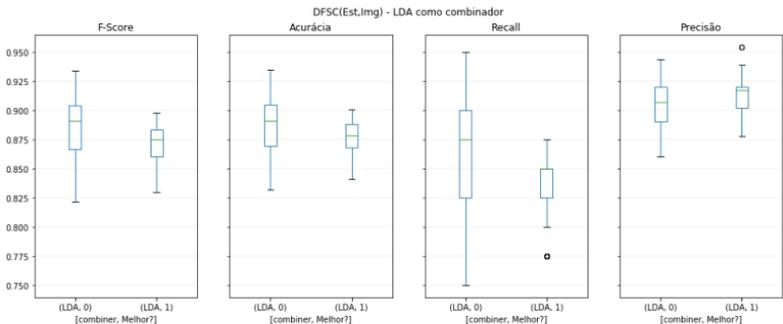
Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.865 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0021, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.963 com p-value: 0.0021 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.823 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.928 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.898 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 33 que:

Figura 33 – DFSC(Est,Img) - Usando LDA como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se:

Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0038, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.966 com p-value: 0.0038 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.896 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

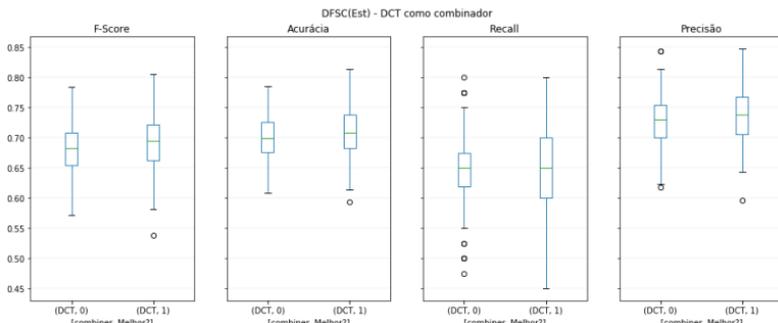
Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0202, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.974 com p-value: 0.0202 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.907 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.935 com p-value: 0.0000 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.825 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0016, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.961 com p-value: 0.0016 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0000, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.898 com p-value: 0.0000 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Pode-se afirmar que houve uma mudança estatisticamente significativa

Utilizando como combinador o algoritmo DCT, pode-se dizer que: Considerando a abordagem DFSC(Est), pode-se observar na Figura 34 que:

Figura 34 – DFSC(Est) - Usando DCT como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.3201, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.987 com p-value: 0.3201 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.1961, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.985 com p-value: 0.1961

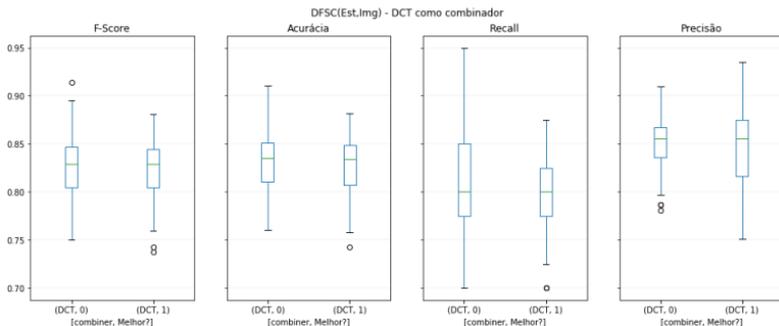
Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.8659, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.994 com p-value: 0.8659 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.6716, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.991 com p-value: 0.6716

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.0119, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.971 com p-value: 0.0119 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.0673, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.980 com p-value: 0.0673

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.4636, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.989 com p-value: 0.4636 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-value é 0.7578, logo podemos dizer que se comporta como uma Distribuição Normal depois 0.992 com p-value: 0.7578

Considerando a abordagem DFSC(Est,Img), pode-se observar na Figura 35 que:

Figura 35 – DFSC(Est,Img) - Usando DCT como combinador - Impacto nas métricas ao considerar os Melhores contra apenas o Melhor



Para a métrica F-Score, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.6830, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.992 com p-value: 0.6830 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0246, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.975 com p-value: 0.0246

Para a métrica Acurácia, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.7976, logo podemos dizer que se comporta como uma Distribuição Normal antes 0.993 com p-value: 0.7976 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0205, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.974 com p-value: 0.0205

Para a métrica Recall, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0106, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.971 com p-value: 0.0106 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0007, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.957 com p-value: 0.0007 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa

Para a métrica Precisão, pode-se dizer que: Aplicando o teste de T-pareado, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa usando o

Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0244, logo podemos dizer que não se comporta como uma Distribuição Normal antes 0.975 com p-value: 0.0244 usando o Teste de Shapiro-Wilk, pode-se dizer que: p-valor é 0.0159, logo podemos dizer que não se comporta como uma Distribuição Normal depois 0.973 com p-value: 0.0159 Ambas não se comportam como uma Distribuição Normal, logo, deve-se aplicar o Teste de Wilcoxon. Aplicando o teste de Wilcoxon, têm-se: Não se pode afirmar que houve uma mudança estatisticamente significativa.

APÊNDICE C – DETALHAMENTO RELEVÂNCIA ESTATÍSTICA DFSC(EST,IMG) PARA CADA UMA DAS MÉTRICAS

Utilizando como combinador o algoritmo GNB, pode-se dizer que:

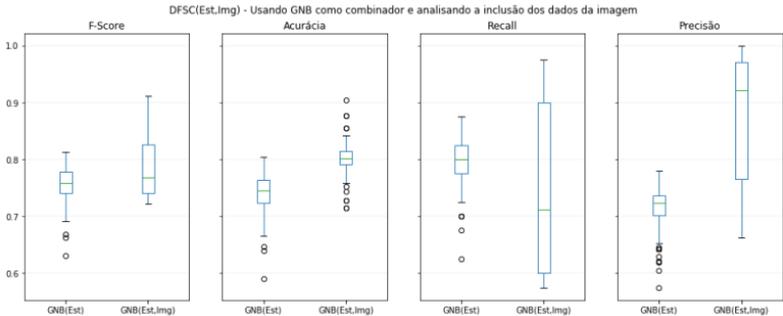


Figura 36 – DFSC(Est,Img) - Usando GNB como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 36, pode-se observar que: Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.920, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=5712.0, pvalue=4.467333938048867e-16) . A partir da análise do p-value é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.953, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.9861096326987145e-41) . A partir da análise do p-value é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.795, p-value: 0.0000). A partir da

análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se $Wilcoxon(statistic=7856.5, pvalue=8.03588751685963e-10)$. A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.834, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se $Wilcoxon(statistic=0.0, pvalue=4.006549346395347e-41)$. A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo KNN, pode-se dizer que:

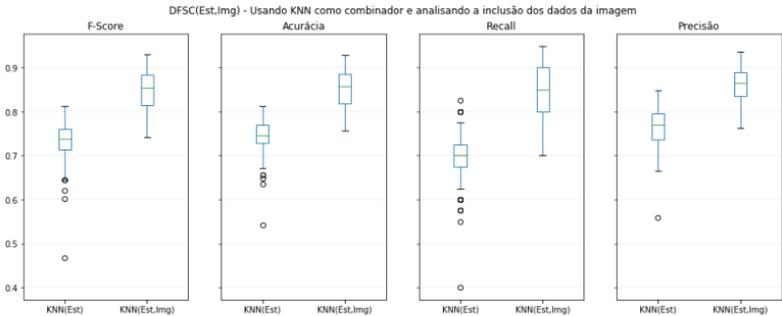


Figura 37 – $DFSC(Est,Img)$ - Usando KNN como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem $DFSC(Est)$ com $DFSC(Est,Img)$ considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 37, pode-se observar que: Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.966, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se $Wilcoxon(statistic=0.0, pvalue=4.0145497654072276e-41)$. A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.972, p-value: 0.0001). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de

Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.006627706137134e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.966, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.1183648034609307e-40) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.934, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.014510509586455e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo RF, pode-se dizer que:

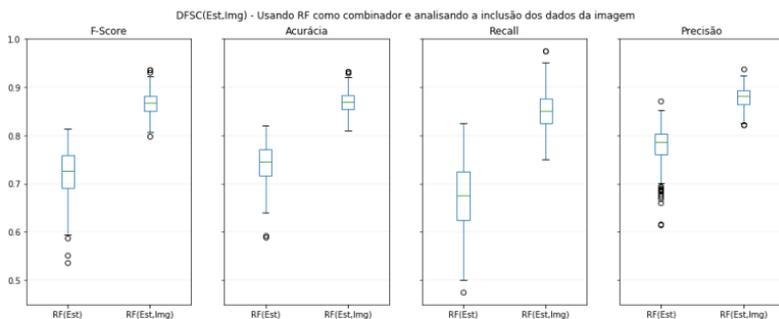


Figura 38 – DFSC(Est,Img) - Usando RF como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 38, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.980, p-value: 0.0015). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.013450744316022e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.978, p-value: 0.0009). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.9980562119845567e-41) . A partir da análise do p-value é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.981, p-value: 0.0029). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.3909176513022246e-41) . A partir da análise do p-value é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.948, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.0130190665407257e-41) . A partir da análise do p-value é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo MLP, pode-se dizer que:

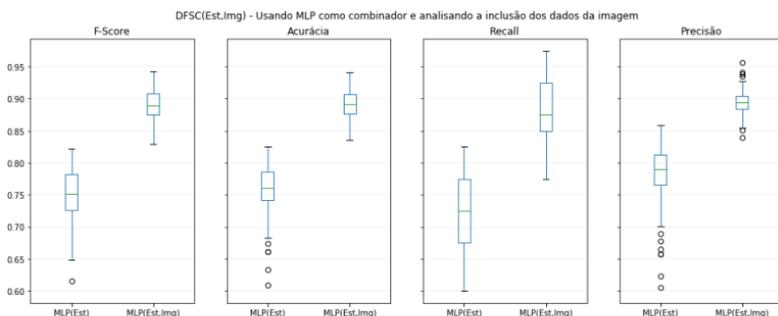


Figura 39 – DFSC(Est,Img) - Usando MLP como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 39, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.993, p-value: 0.3510). A partir da análise do p-value entende-se que o comportamento é comparável a uma Distribuição

Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-69.468, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.983, p-value: 0.0064). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.995085823327709e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.976, p-value: 0.0004). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.4047902712533733e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.929, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.011410485755414e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo LGR, pode-se dizer que:

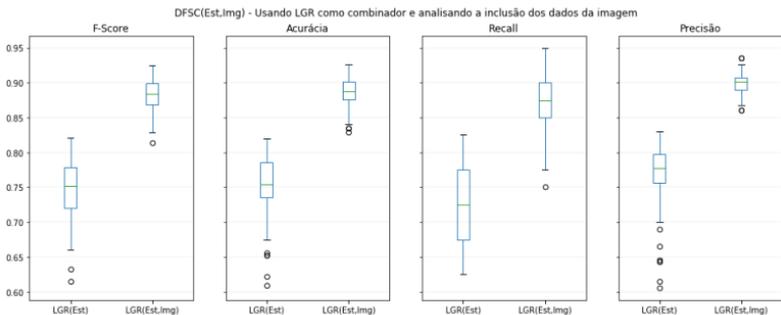


Figura 40 – DFSC(Est,Img) - Usando LGR como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando

apenas a inclusão dos dados da imagem, conforme a Figura 40, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.985, p-value: 0.0144). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.004003470547046e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.974, p-value: 0.0002). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=3.9870843493372117e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.976, p-value: 0.0004). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=2.952375005998577e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.930, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.003377036112874e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo LDA, pode-se dizer que:

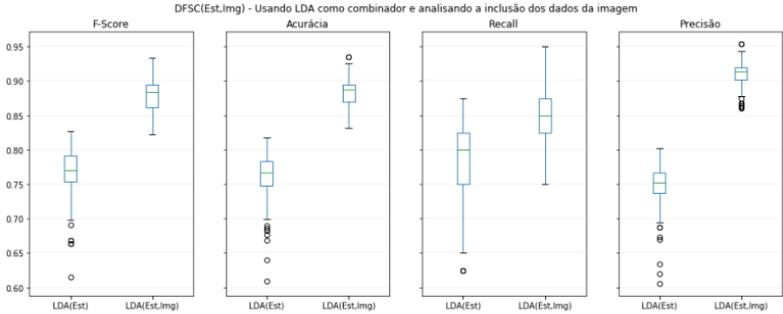


Figura 41 – $DFSC(Est,Img)$ - Usando LDA como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem $DFSC(Est)$ com $DFSC(Est,Img)$ considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 41, pode-se observar que: Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.994, p-value: 0.4155). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-55.394, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.991, p-value: 0.1581). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-65.148, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.971, p-value: 0.0001). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=673.5, pvalue=4.488653913954333e-32) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.974, p-value: 0.0002). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de

Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=4.0076465191845146e-41) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Utilizando como combinador o algoritmo DCT, pode-se dizer que:

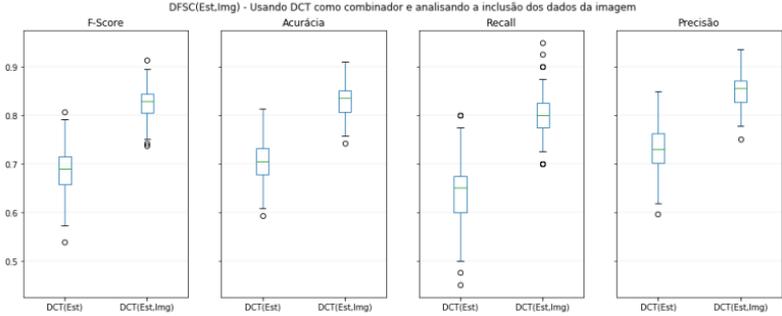


Figura 42 – DFSC(Est,Img) - Usando DCT como combinador e analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem, conforme a Figura 42, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.993, p-value: 0.3218). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-37.057, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.994, p-value: 0.4572). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-38.723, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.976, p-value: 0.0005). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=51.0, pvalue=1.4575705080492432e-40) . A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.996, p-value: 0.8588). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-32.933, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve uma mudança estatisticamente significativa.

APÊNDICE D – ESTUDOS COMPLEMENTARES COM DFSC(EST,IMG)

Análogo ao apresentado no Capítulo 5.8, entretanto foram apresentadas algumas alternativas de utilização do DFSC(Est), que podem ser:

- considerando o melhor modelo de um algoritmo ou considerar os melhores modelos de um algoritmo
- variando a quantidade de algoritmos utilizados

Analogamente, após essas manipulações em DFSC(Est), os resultados foram agregados aos DFSC(Img) e, por meio de um algoritmo combinador, classificou o anúncio, o que configura o DFSC(Est,Img).

D.1 DFSC(Est,Img) - Melhor ou Melhores

D.1.1 Resultado

Na Tabela 25, é possível observar detalhadamente como este comportamento se reflete em cada um dos algoritmos combinadores. Desse modo, é possível que seja observada a abordagem considerando um algoritmo combinador que irá avaliar a saída de DFSC(Est) e combiná-lo com DFSC(Img), contudo, como pode ser observado na tabela, também é apresentado casos onde não se considera a imagem, logo seria um algoritmo combinador que avaliaria apenas o DFSC(Est) e casos onde não são avaliados outros senão apenas o melhor, desse modo em casos onde haja apenas um algoritmo o combinador estaria atuando como um classificador de um dado já previamente classificado, servindo para definição de valor de corte apenas, com base nos dados do conjunto de treino. Porém, para permitir uma ampla comparação essas ressalvas existem e será feito um segundo estudo no D.2, onde foram analisadas as quantidades de algoritmos em DFSC(Est). Além disso são apresentadas na tabela as métricas e seus respectivos desvios padrões quando aplicadas ao conjunto teste (D'_{tt}).

Conforme pode ser observado na Figura 43, a consideração da imagem, por meio de DFSC(Est,Img), aparentemente, traz ganhos independente de combinar DFSC(Est) feito com o melhor de cada algoritmo ou os melhores por algoritmo. Contudo, a seguir, foi feito um estudo quanto à relevância estatística. A sugestão inicial é que o uso da imagem influencia a detecção de fraude.

DFSC(Est) comparado com DFSC(Est,Img) considerando variações na parte DFSC(Est) usando combinações entre o melhor de cada algoritmo ou melhores de cada algoritmo

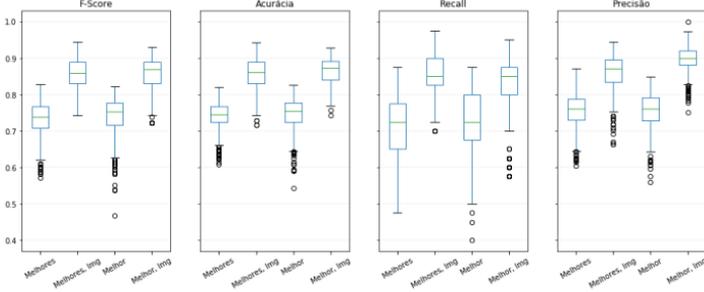


Figura 43 – Visão Geral DFSC(Est,Img)

Tabela 25 – Comparação de Médias \pm Desvio Padrão ($\mu \pm \sigma$) para Algoritmos Combinadores Considerando a Abordagem DFSC(Est,Img) aplicados ao conjunto Teste (D_{tt})

Combinador	Considera Imagem?	Apenas Melhor?	F-Score	Acurácia	Recall	Precisão
DCT	0	0	0.713 \pm 0.0592	0.727 \pm 0.0512	0.683 \pm 0.0765	0.749 \pm 0.0531
DCT	0	1	0.717 \pm 0.0595	0.730 \pm 0.0513	0.691 \pm 0.0780	0.749 \pm 0.0506
DCT	1	0	0.842 \pm 0.0383	0.845 \pm 0.0358	0.831 \pm 0.0542	0.855 \pm 0.0382
DCT	1	1	0.845 \pm 0.0378	0.848 \pm 0.0357	0.829 \pm 0.0526	0.863 \pm 0.0384
GAU	0	0	0.782 \pm 0.0333	0.761 \pm 0.0333	0.862 \pm 0.0628	0.718 \pm 0.0285
GAU	0	1	0.777 \pm 0.0320	0.757 \pm 0.0312	0.852 \pm 0.0605	0.716 \pm 0.0281
GAU	1	0	0.840 \pm 0.0247	0.822 \pm 0.0280	0.935 \pm 0.0421	0.764 \pm 0.0314
GAU	1	1	0.756 \pm 0.0241	0.799 \pm 0.0150	0.625 \pm 0.0346	0.958 \pm 0.0133
KNN	0	0	0.773 \pm 0.0510	0.774 \pm 0.0433	0.777 \pm 0.0828	0.772 \pm 0.0332
KNN	0	1	0.762 \pm 0.0593	0.769 \pm 0.0502	0.745 \pm 0.0860	0.783 \pm 0.0452
KNN	1	0	0.845 \pm 0.0397	0.847 \pm 0.0369	0.839 \pm 0.0592	0.852 \pm 0.0341
KNN	1	1	0.886 \pm 0.0291	0.887 \pm 0.0272	0.884 \pm 0.0460	0.889 \pm 0.0239
LDA	0	0	0.801 \pm 0.0478	0.793 \pm 0.0452	0.838 \pm 0.0743	0.769 \pm 0.0333
LDA	0	1	0.797 \pm 0.0436	0.784 \pm 0.0400	0.853 \pm 0.0783	0.750 \pm 0.0319
LDA	1	0	0.904 \pm 0.0336	0.906 \pm 0.0307	0.898 \pm 0.0589	0.913 \pm 0.0199
LDA	1	1	0.885 \pm 0.0245	0.890 \pm 0.0213	0.851 \pm 0.0422	0.923 \pm 0.0157
LGR	0	0	0.788 \pm 0.0582	0.790 \pm 0.0494	0.793 \pm 0.0951	0.789 \pm 0.0338
LGR	0	1	0.787 \pm 0.0445	0.782 \pm 0.0398	0.810 \pm 0.0750	0.767 \pm 0.0332
LGR	1	0	0.902 \pm 0.0315	0.903 \pm 0.0290	0.896 \pm 0.0500	0.909 \pm 0.0171
LGR	1	1	0.899 \pm 0.0285	0.901 \pm 0.0257	0.889 \pm 0.0495	0.912 \pm 0.0153
MLP	0	0	0.785 \pm 0.0543	0.789 \pm 0.0444	0.780 \pm 0.0924	0.795 \pm 0.0330
MLP	0	1	0.787 \pm 0.0475	0.785 \pm 0.0425	0.802 \pm 0.0779	0.777 \pm 0.0366
MLP	1	0	0.899 \pm 0.0276	0.899 \pm 0.0253	0.897 \pm 0.0481	0.901 \pm 0.0204
MLP	1	1	0.901 \pm 0.0267	0.903 \pm 0.0242	0.895 \pm 0.0477	0.910 \pm 0.0164
RF	0	0	0.775 \pm 0.0688	0.788 \pm 0.0566	0.742 \pm 0.0952	0.815 \pm 0.0421
RF	0	1	0.756 \pm 0.0633	0.767 \pm 0.0555	0.727 \pm 0.0795	0.790 \pm 0.0543
RF	1	0	0.892 \pm 0.0341	0.892 \pm 0.0331	0.892 \pm 0.0448	0.892 \pm 0.0327
RF	1	1	0.883 \pm 0.0275	0.885 \pm 0.0259	0.870 \pm 0.0387	0.897 \pm 0.0240

Para verificar a relevância estatística aplicou-se o Teste-t pareado com o objetivo

de verificar se a média das diferenças entre duas amostras pareadas é diferente de 0. Este foi o teste escolhido, pois é indicado quando se quer comparar o mesmo conjunto de itens após a medição sob duas condições diferentes, DFSC(Est) e DFSC(Est,Img).

Porém, o Teste-t pareado requer que:

- a distribuição seja aproximadamente uma distribuição Normal. Ainda que a distribuição não seja Normal(desde que a variação seja pequena), o Teste-T pareado é robusto à essa violação, podendo ser considerado valido.
- sem *outliers* significativos.

D.1.1.1 Avaliações e Relevância Estatística da inclusão da imagem

Nas Tabelas 26 e 26, podem ser observadas, em resumo, para cada uma das métricas se a inclusão da imagem altera de modo significativo ou não. Na primeira coluna é apresentado qual foi o algoritmo combinador, na segunda coluna são avaliados se foram considerados o melhor de cada um dos algoritmos que foram sujeitos ao combinador ou se foram submissos os melhores de diversos algoritmos, em seguida é expressa a métrica avaliada seguida pelo entendimento se é uma distribuição normal, acompanhada de qual teste de hipótese foi utilizado e por fim se houve relevância estatística. Uma análise permite concluir que para todos os casos foram observadas relevâncias estatísticas, reforçando a hipótese inicial desta dissertação.

Tabela 26 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem - parte 01 / 02

Algo-ritmo	DFSC(Est) com melhor(es) algoritmo(s)	Métrica Avaliada	Teste de Normalidade (Shapiro-Wilk)	Distribuição Normal?	Teste de Hipótese	Valores no Teste de Hipótese	Estatisticamente significativa?
GAU	melhores	F-Score	(0.96, 0.003)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	melhores	Acurácia	(0.90, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	melhores	Recall	(0.74, 0.000)	Não	Wilcoxon	(17.50, 0.000)	Sim
GAU	melhores	Precisão	(0.79, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	melhor	F-Score	(0.90, 0.000)	Não	Wilcoxon	(1647.00, 0.000)	Sim
GAU	melhor	Acurácia	(0.89, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	melhor	Recall	(0.92, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
GAU	melhor	Precisão	(0.93, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhores	F-Score	(0.91, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhores	Acurácia	(0.92, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhores	Recall	(0.90, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhores	Precisão	(0.91, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhor	F-Score	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhor	Acurácia	(0.98, 0.027)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhor	Recall	(0.97, 0.022)	Não	Wilcoxon	(0.00, 0.000)	Sim
KNN	melhor	Precisão	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhores	F-Score	(0.99, 0.647)	Sim	T-pareado	(-37.32, 0.000)	Sim
RF	melhores	Acurácia	(0.99, 0.772)	Sim	T-pareado	(-39.46, 0.000)	Sim
RF	melhores	Recall	(0.97, 0.010)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhores	Precisão	(0.97, 0.016)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhor	F-Score	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhor	Acurácia	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhor	Recall	(0.97, 0.022)	Não	Wilcoxon	(0.00, 0.000)	Sim
RF	melhor	Precisão	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim

Tabela 27 – DFSC(Est) comparado DFSC(Est,Img) Relevância Estatística dos Algoritmos Avaliados após inclusão da Imagem - parte 02 / 02

Algo- ritmo	DFSC(Est) com DFSC(Est,Img) algoritmo(s)	Métrica Avaliada	Teste de Normalidade (Shapiro-Wilk)	Distri- buição Normal?	Teste de Hipótese	Valores no Teste de Hipótese	Estatisticamente significativa?
MLP	melhores	F-Score	(0.99, 0.434)	Sim	T-pareado	(-50.37, 0.000)	Sim
MLP	melhores	Acurácia	(0.99, 0.480)	Sim	T-pareado	(-54.00, 0.000)	Sim
MLP	melhores	Recall	(0.97, 0.013)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	melhores	Precisão	(0.93, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	melhor	F-Score	(0.98, 0.091)	Sim	T-pareado	(-49.60, 0.000)	Sim
MLP	melhor	Acurácia	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	melhor	Recall	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
MLP	melhor	Precisão	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhores	F-Score	(0.98, 0.155)	Sim	T-pareado	(-51.34, 0.000)	Sim
LGR	melhores	Acurácia	(0.98, 0.111)	Sim	T-pareado	(-55.41, 0.000)	Sim
LGR	melhores	Recall	(0.97, 0.004)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhores	Precisão	(0.91, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhor	F-Score	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhor	Acurácia	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhor	Recall	(0.96, 0.001)	Não	Wilcoxon	(0.00, 0.000)	Sim
LGR	melhor	Precisão	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	melhores	F-Score	(0.97, 0.003)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	melhores	Acurácia	(0.98, 0.058)	Sim	T-pareado	(-54.72, 0.000)	Sim
LDA	melhores	Recall	(0.89, 0.000)	Não	Wilcoxon	(189.00, 0.000)	Sim
LDA	melhores	Precisão	(0.97, 0.010)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	melhor	F-Score	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	melhor	Acurácia	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
LDA	melhor	Recall	(0.92, 0.000)	Não	Wilcoxon	(66.00, 0.000)	Sim
LDA	melhor	Precisão	(0.95, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
DCT	melhores	F-Score	(0.97, 0.004)	Não	Wilcoxon	(0.00, 0.000)	Sim
DCT	melhores	Acurácia	(0.98, 0.058)	Sim	T-pareado	(-31.18, 0.000)	Sim
DCT	melhores	Recall	(0.94, 0.000)	Não	Wilcoxon	(0.00, 0.000)	Sim
DCT	melhores	Precisão	(0.99, 0.445)	Sim	T-pareado	(-29.73, 0.000)	Sim
DCT	melhor	F-Score	(0.99, 0.778)	Sim	T-pareado	(-24.12, 0.000)	Sim
DCT	melhor	Acurácia	(0.99, 0.629)	Sim	T-pareado	(-24.49, 0.000)	Sim
DCT	melhor	Recall	(0.99, 0.444)	Sim	T-pareado	(-21.18, 0.000)	Sim
DCT	melhor	Precisão	(0.99, 0.931)	Sim	T-pareado	(-19.33, 0.000)	Sim

Contudo, maiores detalhes sobre a análise estatística foram apresentados a seguir.

D.1.2 Detalhamento Relevância Estatística DFSC(Est,Img) para cada uma das métricas

Utilizando como combinador o algoritmo GNB, pode-se dizer que:

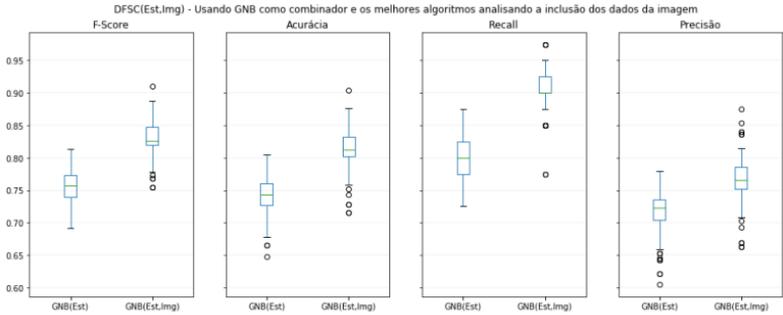


Figura 44 – DFSC(Est,Img) - Usando GNB como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura 44, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.965, p-value: 0.0031). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.964705890962901e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.895, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9318835496631486e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.745, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição

Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=17.5, pvalue=1.499657416000026e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.791, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.964705890962901e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

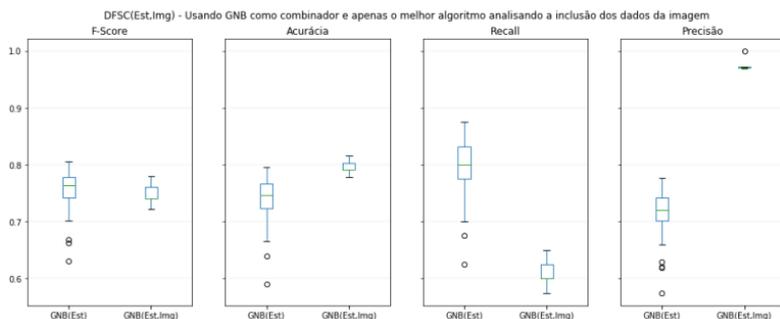


Figura 45 – DFSC(Est,Img) - Usando GNB como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 45, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.904, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=1647.0, pvalue=2.0641874332277187e-07) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.895, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9549567582631292e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.918, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.5108872737199344e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.930, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.962091338594042e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo KNN, pode-se dizer que:

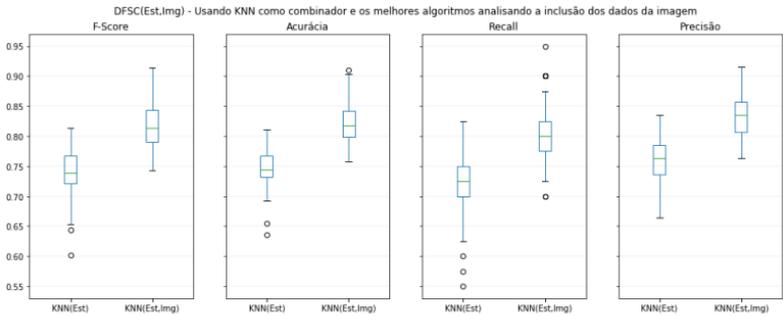


Figura 46 – DFSC(Est,Img) - Usando KNN como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura 46, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.907, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.971411250141284e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.921, p-value: 0.0000). A

partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9635520014751498e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.902, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=5.291025430830579e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.906, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.971411250141284e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

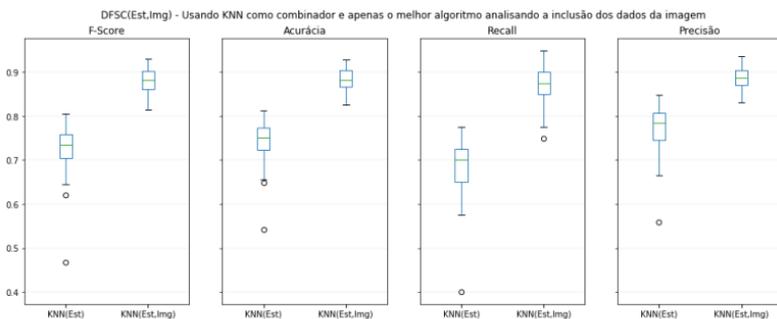


Figura 47 – DFSC(Est,Img) - Usando KNN como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 47, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.955, p-value: 0.0005). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon,

obtem-se Wilcoxon(statistic=0.0, pvalue=1.9712568584476237e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.975, p-value: 0.0269). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.968171452047417e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.974, p-value: 0.0220). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.788241203835595e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.957, p-value: 0.0008). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.971179666937574e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo RF, pode-se dizer que:

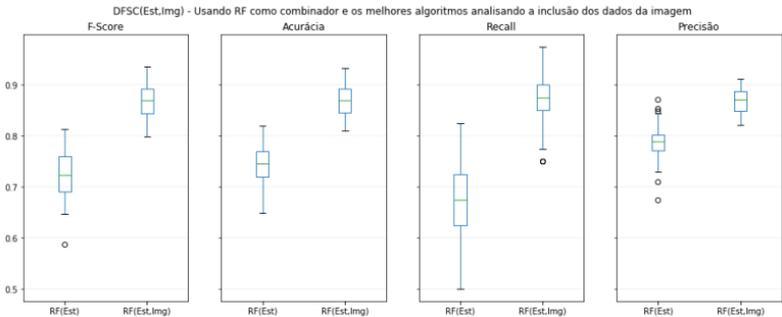


Figura 48 – DFSC(Est,Img) - Usando RF como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura

48, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.991, p-value: 0.6472). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-37.318, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.993, p-value: 0.7720). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-39.462, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.971, p-value: 0.0099). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.7708123619185137e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.973, p-value: 0.0159). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9704079108279912e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

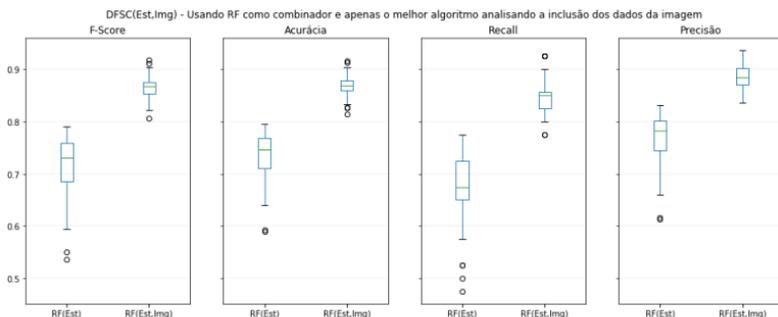


Figura 49 – DFSC(Est,Img) - Usando RF como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 49, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.957, p-value: 0.0007). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9710252925905593e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.959, p-value: 0.0011). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.96439812370122e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.974, p-value: 0.0219). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.7581069296841126e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.944, p-value: 0.0001). A

partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se $Wilcoxon(statistic=0.0, pvalue=1.97071657858581e-21)$. A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo MLP, pode-se dizer que:

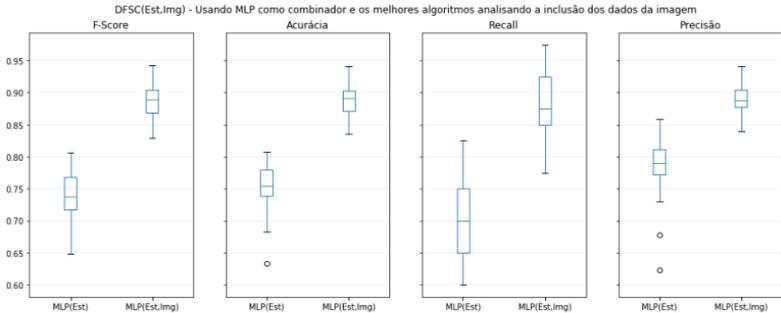


Figura 50 – DFSC(Est,Img) - Usando MLP como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura 50, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.989, p-value: 0.4343). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-50.370, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.989, p-value: 0.4795). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-54.003, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.972, p-value: 0.0126). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon,

obtem-se Wilcoxon(statistic=0.0, pvalue=1.7943693885136853e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.926, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.971488450325114e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

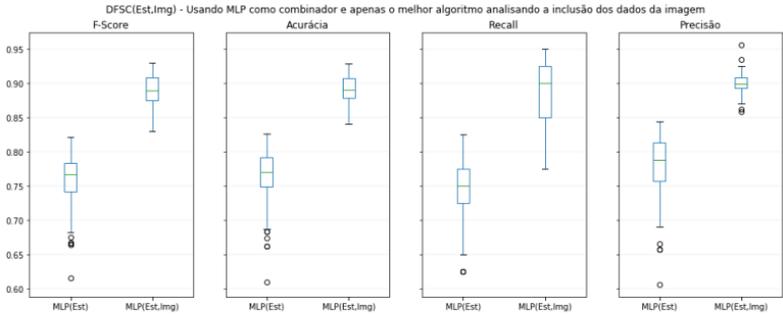


Figura 51 – DFSC(Est,Img) - Usando MLP como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 51, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.981, p-value: 0.0906). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-49.603, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.959, p-value: 0.0010). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9636289072711687e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.957, p-value: 0.0007). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.7474137493882874e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.943, p-value: 0.0001). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.967477872432308e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo LGR, pode-se dizer que:

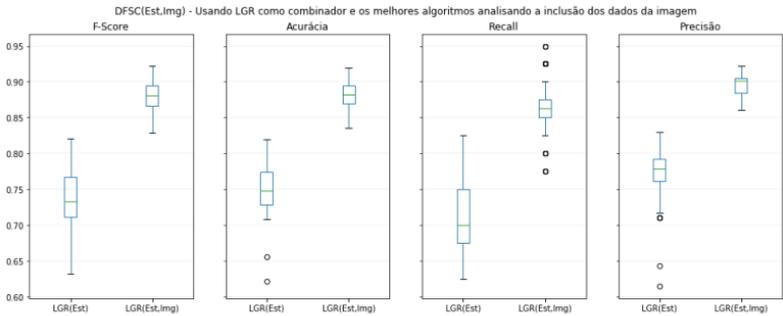


Figura 52 – DFSC(Est,Img) - Usando LGR como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura 52, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.984, p-value: 0.1546). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-51.339, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.982, p-value: 0.1108).

A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-55.408, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.966, p-value: 0.0043). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.733844546561067e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.906, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9660144164119998e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

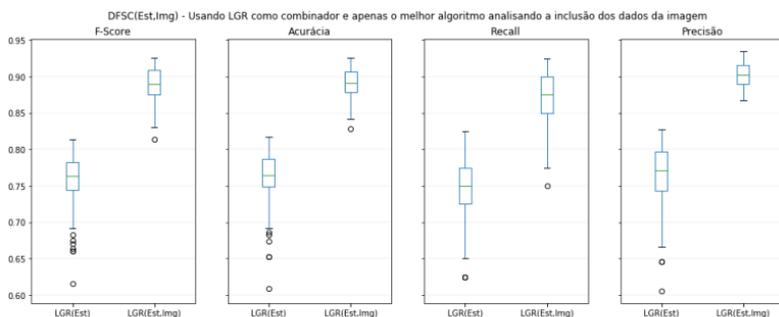


Figura 53 – DFSC(Est,Img) - Usando LGR como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 53, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.955, p-value: 0.0005). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon,

obtem-se Wilcoxon(statistic=0.0, pvalue=1.9592498717219984e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.945, p-value: 0.0001). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9502141636886e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.960, p-value: 0.0014). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.4448545026433122e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.947, p-value: 0.0001). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9592498717219984e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo LDA, pode-se dizer que:

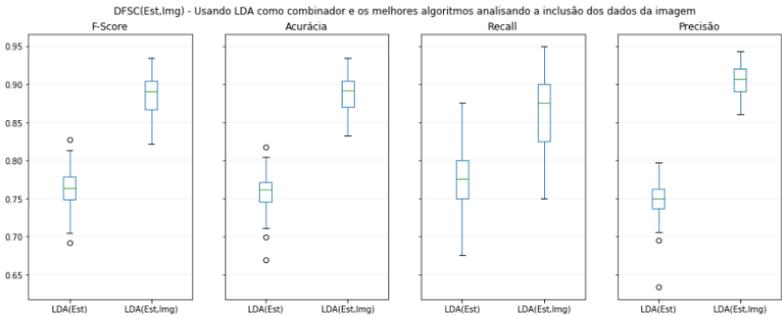


Figura 54 – DFSC(Est,Img) - Usando LDA como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura

54, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.965, p-value: 0.0035). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.969790714026798e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.979, p-value: 0.0584). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-54.717, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.889, p-value: 0.0000). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=189.0, pvalue=3.656564706122192e-19) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.971, p-value: 0.0105). A partir da análise do p-valor entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.969790714026798e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

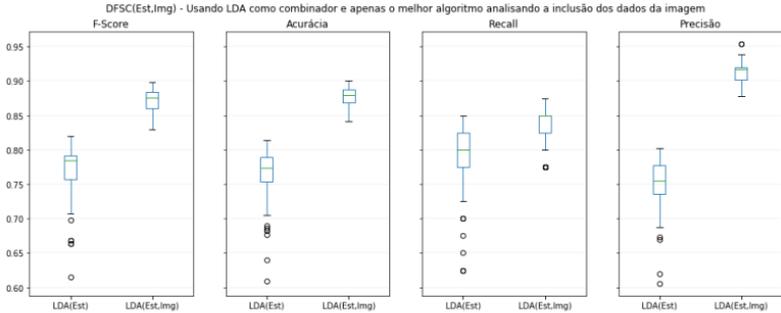


Figura 55 – DFSC(Est,Img) - Usando LDA como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 55, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.936, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.961092536528338e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.951, p-value: 0.0002). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9600942207700948e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.918, p-value: 0.0000). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=66.0, pvalue=7.8526414129864e-16) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.951, p-value: 0.0002). A partir da análise do p-value entende-se que o comportamento não é comparável a uma

Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9610925365283383e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Utilizando como combinador o algoritmo DCT, pode-se dizer que:

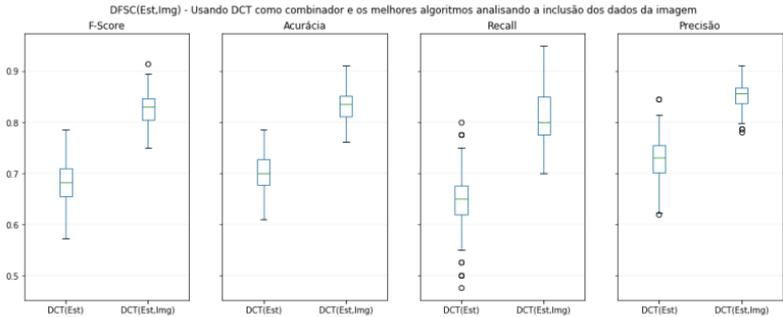


Figura 56 – DFSC(Est,Img) - Usando DCT como combinador e os melhores algoritmos analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando os melhores algoritmos e variando apenas a inclusão dos dados da imagem, conforme a Figura 56, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.966, p-value: 0.0037). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon, obtém-se Wilcoxon(statistic=0.0, pvalue=1.9710252925905593e-21) . A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.979, p-value: 0.0575). A partir da análise do p-value entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-31.181, p-value: 0.0000). A partir da análise do p-value é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.943, p-value: 0.0001). A partir da análise do p-value entende-se que o comportamento não é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste de Wilcoxon. Com o Teste de Wilcoxon,

obtm-se Wilcoxon(statistic=0.0, pvalue=1.850071174167992e-21) . A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.989, p-value: 0.4448). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-29.728, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

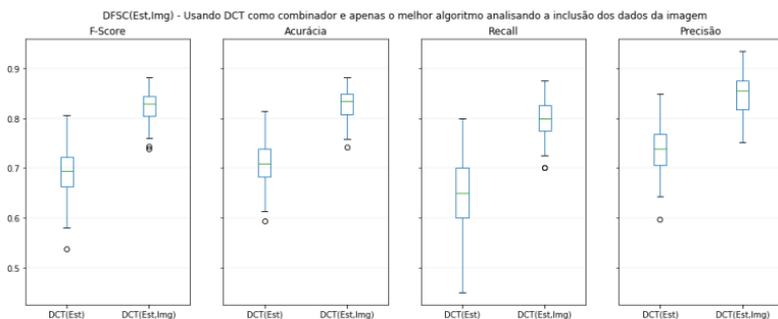


Figura 57 – DFSC(Est,Img) - Usando DCT como combinador e apenas o melhor algoritmo analisando a inclusão dos dados da imagem

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando apenas o melhor algoritmo e variando apenas a inclusão dos dados da imagem, conforme a Figura 57, pode-se observar que:

Para a métrica F-Score, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.993, p-value: 0.7781). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-24.125, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Acurácia, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.991, p-value: 0.6291). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-24.492, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Recall, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.989, p-value: 0.4438). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-21.178, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

Para a métrica Precisão, pode-se dizer que: Com a aplicação do Teste de Shapiro-Wilk, observa-se que a distribuição da diferença tem valores (0.995, p-value: 0.9306). A partir da análise do p-valor entende-se que o comportamento é comparável a uma Distribuição Normal. Desse modo, é necessário aplicar o Teste T-pareado. Com o Teste T-pareado, obtém-se (-19.334, p-value: 0.0000). A partir da análise do p-valor é possível afirmar que houve mudança estatisticamente significativa.

D.2 DFSC(Est,Img) - Variando a quantidade de modelos

D.2.1 Resultado

Conforme pode ser observado na Figura 58, a consideração da imagem, por meio de DFSC(Est,Img), aparentemente, traz ganhos independente de quantidade de algoritmos DFSC(Est) desse modo não será analisada a contribuição do melhor de cada algoritmo comparado com os melhores por algoritmo.

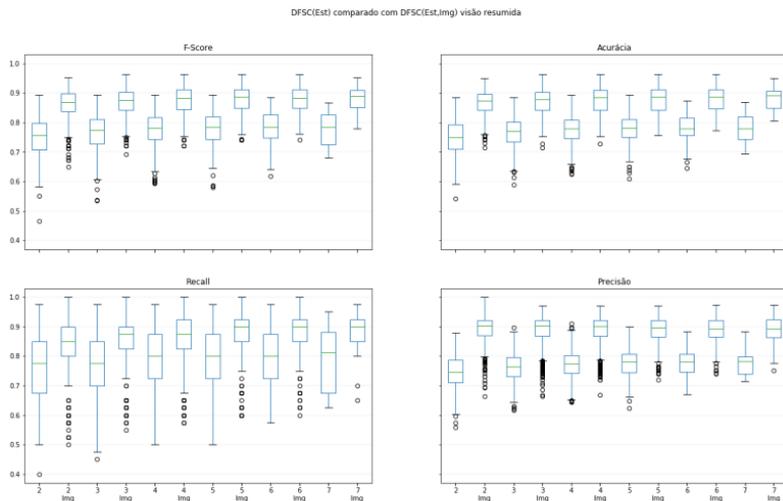


Figura 58 – Visão Geral DFSC(Est,Img) considerando a partir de 2 modelos

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura ??, observa-se que: Utilizando como combinador o algoritmo KNN, pode-se dizer que a utilização de imagens trouxe maior robustez para todas as métricas independente da quantidade de algoritmos Est utilizados.

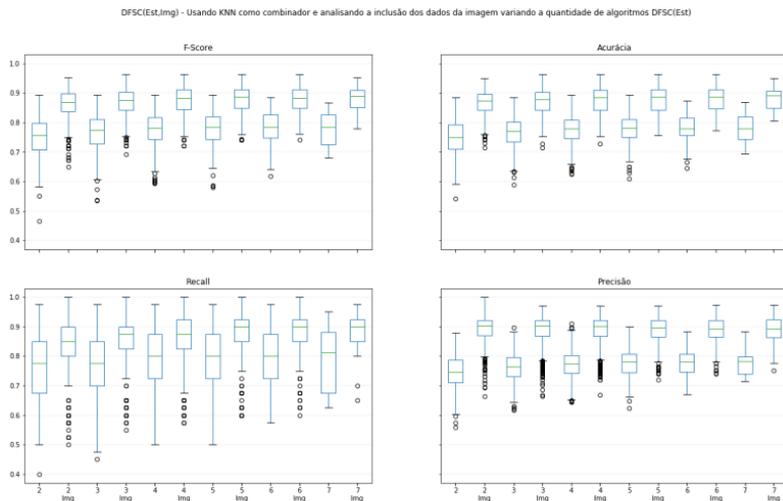


Figura 59 – DFSC(Est,Img) - Usando KNN como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura 59, observa-se que: Utilizando como combinador o algoritmo RF, pode-se dizer que a utilização de imagens trouxe maior robustez para todas as métricas independente da quantidade de algoritmos Est utilizados.

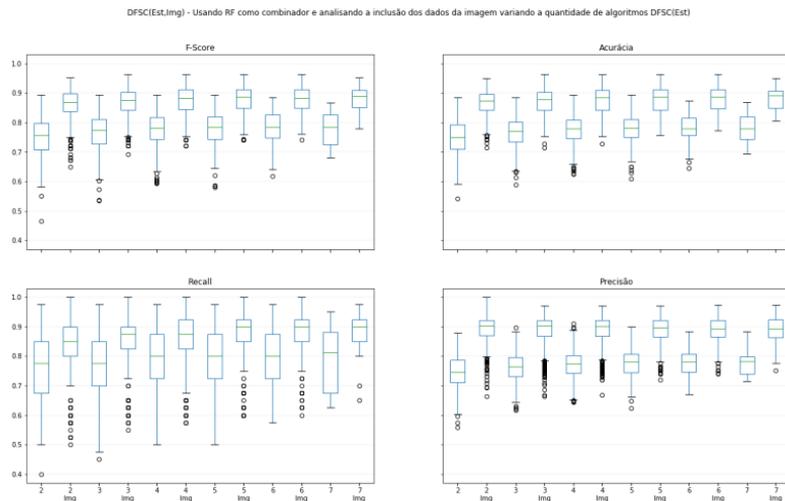


Figura 60 – DFSC(Est,Img) - Usando RF como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura 60, observa-se que: Utilizando como combinador o algoritmo MLP, pode-se dizer que:

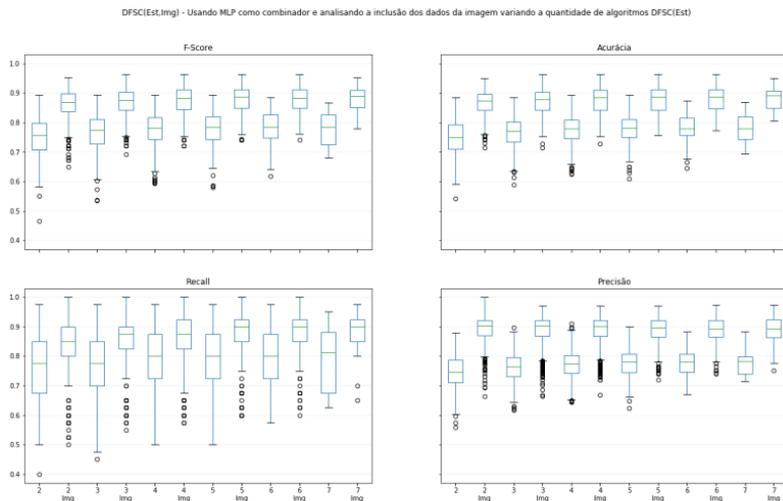


Figura 61 – DFSC(Est,Img) - Usando MLP como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura 61, observa-se que: Utilizando como combinador o algoritmo LGR, pode-se dizer que a utilização de imagens trouxe maior robustez para todas as métricas independente da quantidade de algoritmos Est utilizados.

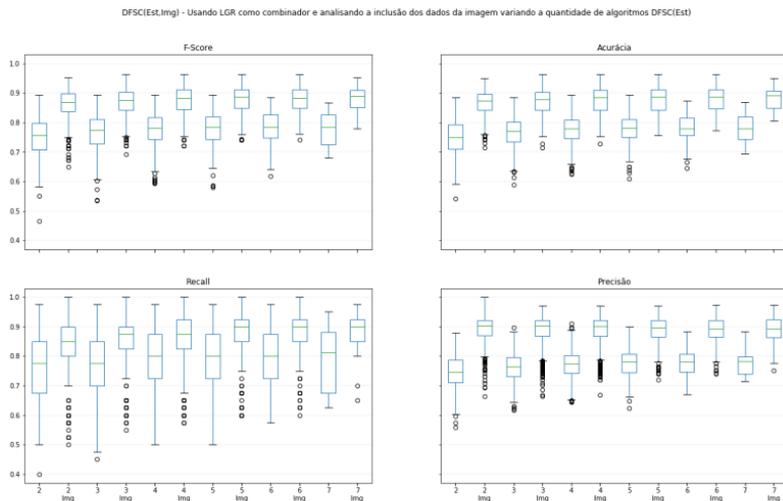


Figura 62 – DFSC(Est,Img) - Usando LGR como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura 62, observa-se que: Utilizando como combinador o algoritmo LDA, pode-se dizer que a utilização de imagens trouxe maior robustez para todas as métricas independente da quantidade de algoritmos Est utilizados.

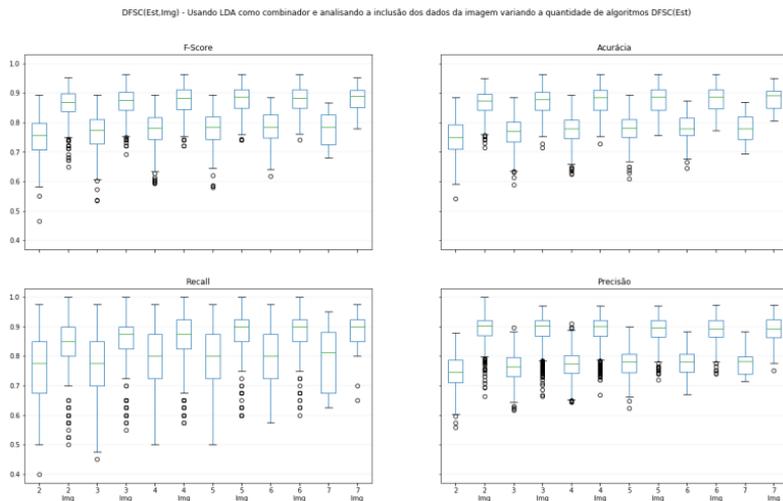


Figura 63 – DFSC(Est,Img) - Usando LDA como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)

Comparando a abordagem DFSC(Est) com DFSC(Est,Img) considerando e variando apenas a inclusão dos dados da imagem ao longo de cada uma das contagens de modelos DFSC(Est), conforme a Figura 63, observa-se que: Utilizando como combinador o algoritmo DCT, pode-se dizer que a utilização de imagens trouxe maior robustez para todas as métricas independente da quantidade de algoritmos Est utilizados.

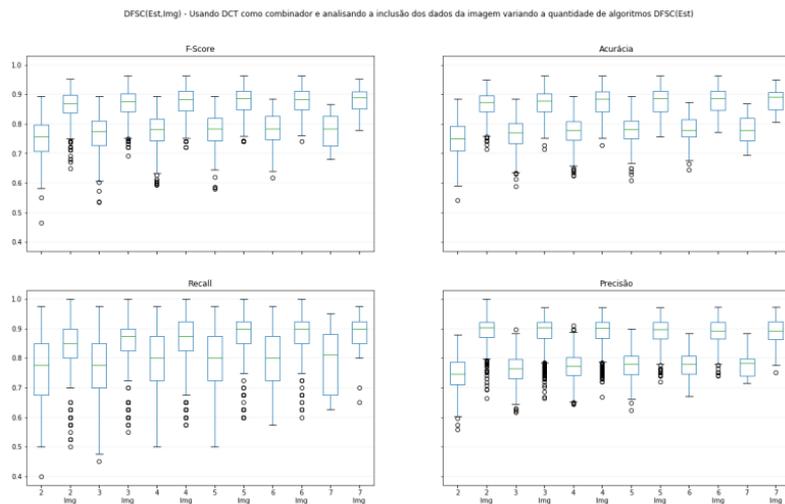


Figura 64 – DFSC(Est,Img) - Usando DCT como combinador e analisando a inclusão dos dados da imagem variando a quantidade de algoritmos DFSC(Est)