

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas – ICEX
Departamento de Estatística

Fernando Augusto Melo Duarte Malta

**Redes Neurais para Dados Tabulares: Uma Comparação
Empírica**

Belo Horizonte

2022

Fernando Augusto Melo Duarte Malta

Redes Neurais para Dados Tabulares: Uma Comparação Empírica

Dissertação apresentada ao Programa de Pós
Graduação em Estatística da Universidade Fe-
deral de Minas Gerais, como requisito para a
obtenção do grau de Mestre em Estatística

Orientador: Dr. Luiz Henrique Duczmal

Belo Horizonte
2022

Malta, Fernando Augusto Melo Duarte.

M261r Redes neurais para dados tabulares[manuscrito]: uma
comparação empírica / Fernando Augusto Melo Duarte Malta –
2022.
124 f. il.

Orientador: Luiz Henrique Duczmal.
Dissertação (mestrado) - Universidade Federal de Minas
Gerais, Instituto de Ciências Exatas, Departamento de Estatística
Referências: f.121-124.

1. Estatística – Teses. 2. Redes neurais – Teses. 3. Predição
– Teses. 4. Gradient Boosted Trees – Teses. I. Duczmal, Luiz
Henrique. II. Universidade Federal de Minas Gerais, Instituto de
Ciências Exatas, Departamento de Estatística. III. Título.

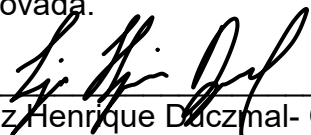
CDU 519.2(043)



ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO DO FERNANDO AUGUSTO MELO DUARTE MALTA, MATRICULADO, SOB O Nº 2020.666.949, NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 25 DE MAIO DE 2022.


Aos 25 dias do mês de Maio de 2022, às 14h00, em reunião pública virtual 266 (conforme orientações para a atividade de defesa de dissertação durante a vigência da Portaria PRPG nº 1819) OU na sala 2076 do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de dissertação do aluno FERNANDO AUGUSTO MELO DUARTE MALTA, nº matrícula 2020666949, intitulada: “ *Uma Comparação Empírica* ”, requisito final para obtenção do Grau de mestre em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Luiz Henrique Duczmal, passou a palavra ao aluno para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do aluno. Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença do aluno e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

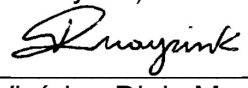
- (x) Aprovada.
- () Reprovada com resubmissão do texto em ____ dias.
- () Reprovada com resubmissão do texto e nova defesa em ____ dias.
- () Reprovada.

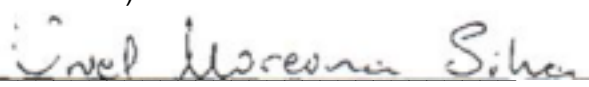

Prof. Luiz Henrique Duczmal- Orientador
(EST/UFMG)

MILTON PIFANO SOARES FERREIRA:52368270604
Assinado de forma digital por MILTON PIFANO SOARES FERREIRA:52368270604
Dados: 2022.05.31 09:08:25 -03'00'

Milton Pifano Soares Ferreira
Co-Orientador (Fundação Dom Cabral – Hence Analytics)


Profa. Denise Bulgarelli Duczmal
(MAT/UFMG)


Prof. Vinícius Diniz Mayrink
(EST/UFMG)


Prof. Uriel Moreira Silva (Doutor/EST/UFMG)

O resultado final foi comunicado publicamente ao aluno pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 25 de maio de 2022.

Observações:

1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá pela aprovação ou reprovação da tese.

AGRADECIMENTOS

À minha mãe, Iara, que sempre esteve ao meu lado; a Deus; ao meu melhor amigo, Hugo; ao meu orientador, professor Luiz; ao Uriel e ao Milton. Me faltam palavras para expressar tamanha gratidão por todos que me apoiaram nessa jornada.

*And the cat's in the cradle and the silver spoon
Little boy blue and the man in the moon
"Cat's in the Cradle" - Henry Chapin*

RESUMO

O Aprendizado Profundo passou por significativas transformações nas últimas décadas, o tornando a ferramenta principal na modelagem de dados não estruturados, como vídeos, áudios, linguagem e imagens. Entretanto, para dados estruturados ele não foi capaz de superar a popularidade de métodos mais estabelecidos, como Modelos Lineares Generalizados, *Gradient Boosted Trees* e *Bagging*.

Nos últimos anos, modelos promissores emergiram para dados tabulares objetivando adaptar, para redes neurais, inovações aplicadas à dados não estruturados, como a arquitetura dos *transformers*, ou que tentam simular processos baseados em árvores de decisão.

Este estudo é uma comparação empírica desses modelos em bases de dados de dimensões e tamanhos amostrais muito diferentes. Sendo ilustrado por diversas áreas, como ciência dos materiais, marketing, biologia e astronomia.

Palavras-chave: Dados Tabulares, Redes Neurais, Modelagem Preditiva, Gradient Boosted Trees, Estatística

ABSTRACT

Deep Learning has undergone significant transformations in recent decades, making it the dominant strategy in modeling unstructured data, such as videos, audios, language, and images. Although it came to structured data, it was not able to overcome the popularity of more established methods such as Generalized Linear Models, Gradient Boosted Trees and Bagging.

In recent years, promising models have emerged for tabular data that aim to adapt, for neural networks, innovations applied to unstructured data, such as the architecture of transformers, or that tried to simulate processes based on decision trees.

This study is an empirical comparison of these models in databases of very different dimensions and sample sizes. It is illustrated by various areas, such as materials science, marketing, biology, and astronomy.

Keywords: Tabular Data, Neural Networks, Predictive Analytics, Gradient Boosted Trees, Statistics

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação de Árvores de Decisão	22
Figura 2 – Representação do Processo de Bagging	23
Figura 3 – Representação do Processo de Boosting	24
Figura 4 – Representação das possíveis separações de classes pelo SVM	25
Figura 5 – SVM com um Kernel Polinomial de Grau 3 (esquerda). Uso de um Kernel Radial (direita)	26
Figura 6 – Neurônios Artificiais Computando Operadores Lógicos	26
Figura 7 – Modelo do TLU	27
Figura 8 – Representação de um Perceptron de Múltiplas Camadas com a Adição de Neurônios de Viés	28
Figura 9 – Comparação da Arquitetura de uma Rede Neural Convencional (esquerda) com a de uma Residual Network (direita)	29
Figura 10 – Arquitetura do TabNet	30
Figura 11 – Representação do Mecanismo de Self-Attention para Processamento de Linguagem Natural	31
Figura 12 – Comparação da arquitetura de Self-Attention com a de outras redes neurais	32
Figura 13 – Representação de uma Camada de Multi-Head Attention	33
Figura 14 – Arquitetura do Transformer	34
Figura 15 – Arquitetura do TabTransformer	35
Figura 16 – Feature Tokenizer	36
Figura 17 – Arquitetura do FT-Transformer	36
Figura 18 – Arquitetura do FastFormer	37
Figura 19 – Arquitetura do SAINT	38
Figura 20 – Representação do Mecanismo de Atenção Entre Amostras	38
Figura 21 – Representação da Influência Individual de Variáveis Descrita pelo SHAP	40
Figura 22 – Matriz do Teste de McNemar	41
Figura 23 – Histogramas das Variáveis para a Base de Dados dos Atrasos no SUS	51
Figura 24 – Histogramas das Variáveis para a Base de Dados dos Atrasos no SUS	52
Figura 25 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Atrasos no SUS	52
Figura 26 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Atrasos no SUS	53
Figura 27 – Histogramas das Variáveis Numéricas para a Base de Dados dos Diamantes	56
Figura 28 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Diamantes	57
Figura 29 – Histogramas das Variáveis para a Base de Dados dos Múons	61
Figura 30 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Múons	62
Figura 31 – Histogramas das Variáveis para a Base de Dados dos Supercondutores	66
Figura 32 – Histogramas das Variáveis para a Base de Dados dos Supercondutores	67
Figura 33 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Supercondutores	67
Figura 34 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Supercondutores	68

Figura 35 – Histogramas das Variáveis para a Base de Dados das Pulsares	72
Figura 36 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados das Pulsares	73
Figura 37 – Histogramas das Variáveis para a Base de Dados dos Feijões	76
Figura 38 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Marketing	80
Figura 39 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Marketing	81
Figura 40 – Histogramas das Variáveis para a Base de Dados dos Vinhos	84
Figura 41 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Vinhos	85
Figura 42 – Histogramas das Variáveis para a Base de Dados dos Abalones	89
Figura 43 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Abalones	90
Figura 44 – Histograma das Variáveis para a Base de Dados da Cardiotocografia	94
Figura 45 – Histogramas das Variáveis para a Base de Dados do Concreto	98
Figura 46 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Concreto	99
Figura 47 – Histogramas das Variáveis para a Base de Dados de Bioconcentração	103
Figura 48 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados de Bioconcentração	104
Figura 49 – Histogramas das Variáveis para a Base de Dados do Bóson de Higgs	107

LISTA DE TABELAS

Tabela 1 – Informações dos Bancos de Dados Após o Pré-Processamento	45
Tabela 2 – Métricas para a Base de Dados dos Atrasos no SUS	53
Tabela 3 – SHAP Médio para a Base de Dados dos Atrasos no SUS	54
Tabela 4 – SHAP Absolutos para a Base de Dados dos Atrasos no SUS	54
Tabela 5 – Feature Permutation para a Base de Dados dos Atrasos no SUS	55
Tabela 6 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Atrasos no SUS	55
Tabela 7 – Métricas para a Base de Dados dos Diamantes	58
Tabela 8 – SHAP Médio para a Base de Dados dos Diamantes	58
Tabela 9 – SHAP Absolutos para a Base de Dados dos Diamantes	59
Tabela 10 – Feature Permutation para a Base de Dados dos Diamantes	59
Tabela 11 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Diamantes	60
Tabela 12 – Métricas para a Base de Dados dos Múons	63
Tabela 13 – SHAP Médio para a Base de Dados dos Múons	63
Tabela 14 – SHAP Absolutos para a Base de Dados dos Múons	64
Tabela 15 – Feature Permutation para a Base de Dados dos Múons	64
Tabela 16 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Múons	65
Tabela 17 – Métricas para a Base de Dados dos Supercondutores	69
Tabela 18 – SHAP Médio para a Base de Dados dos Supercondutores	70
Tabela 19 – SHAP Absolutos para a Base de Dados dos Supercondutores	70
Tabela 20 – Feature Permutation para a Base de Dados dos Supercondutores	71
Tabela 21 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Supercondutores	71
Tabela 22 – Métricas para a Base de Dados das Pulsares	73
Tabela 23 – SHAP Médio para a Base de Dados das Pulsares	74
Tabela 24 – SHAP Absolutos para a Base de Dados das Pulsares	74
Tabela 25 – Feature Permutation para a Base de Dados das Pulsares	75
Tabela 26 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados das Pulsares	75
Tabela 27 – Métricas para a Base de Dados dos Feijões	77
Tabela 28 – SHAP Médio para a Base de Dados dos Feijões	78
Tabela 29 – SHAP Absolutos para a Base de Dados dos Feijões	78
Tabela 30 – Feature Permutation para a Base de Dados dos Feijões	79
Tabela 31 – Importância das Variáveis Internamente nos Modelos Lineares para a Base de Dados dos Feijões	79
Tabela 32 – Importância das Variáveis Internamente para a Base de Dados dos Feijões	79
Tabela 33 – Métricas para a Base de Dados do Marketing	81
Tabela 34 – SHAP Médio para a Base de Dados do Marketing	82
Tabela 35 – SHAP Absolutos para a Base de Dados do Marketing	82
Tabela 36 – Feature Permutation para a Base de Dados do Marketing	83
Tabela 37 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Marketing	83

Tabela 38 – Métricas para a Base de Dados dos Vinhos	86
Tabela 39 – SHAP Médio para a Base de Dados dos Vinhos	86
Tabela 40 – SHAP Absolutos para a Base de Dados dos Vinhos	87
Tabela 41 – Feature Permutation para a Base de Dados dos Vinhos	87
Tabela 42 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Vinhos	88
Tabela 43 – Métricas para a Base de Dados dos Abalones	91
Tabela 44 – SHAP Médio para a Base de Dados dos Abalones	92
Tabela 45 – SHAP Absolutos para a Base de Dados dos Abalones	92
Tabela 46 – Feature Permutation para a Base de Dados dos Abalones	93
Tabela 47 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Abalones	93
Tabela 48 – Métricas para a Base de Dados da Cardiotocografia	95
Tabela 49 – SHAP Médio para a Base de Dados da Cardiotocografia	95
Tabela 50 – SHAP Absolutos para a Base de Dados da Cardiotocografia	96
Tabela 51 – Feature Permutation para a Base de Dados da Cardiotocografia	96
Tabela 52 – Importância das Variáveis Internamente nos Modelos Lineares para a Base de Dados da Cardiotocografia	96
Tabela 53 – Importância das Variáveis Internamente para a Base de Dados da Cardiotocografia	97
Tabela 54 – Métricas para a Base de Dados do Concreto	100
Tabela 55 – SHAP Médio para a Base de Dados do Concreto	100
Tabela 56 – SHAP Absolutos para a Base de Dados do Concreto	101
Tabela 57 – Feature Permutation para a Base de Dados do Concreto	102
Tabela 58 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Concreto	102
Tabela 59 – Métricas para a Base de Dados de Bioconcentração	104
Tabela 60 – SHAP Médio para a Base de Dados de Bioconcentração	105
Tabela 61 – SHAP Absolutos para a Base de Dados de Bioconcentração	105
Tabela 62 – Feature Permutation para a Base de Dados de Bioconcentração	106
Tabela 63 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados de Bioconcentração	106
Tabela 64 – Métricas para a Base de Dados do Bóson de Higgs	108
Tabela 65 – SHAP Médio para a Base de Dados do Bóson de Higgs	108
Tabela 66 – SHAP Absolutos para a Base de Dados do Bóson de Higgs	109
Tabela 67 – Feature Permutation para a Base de Dados do Bóson de Higgs	109
Tabela 68 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Bóson de Higgs	110
Tabela 69 – Testes de Hipótese para a Base de Dados dos Atrasos no SUS	111
Tabela 70 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados dos Atrasos no SUS	111
Tabela 71 – Testes de Hipótese para a Base de Dados das Pulsares	112
Tabela 72 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados das Pulsares	112
Tabela 73 – Testes de Hipótese para a Base de Dados do Bóson de Higgs	113
Tabela 74 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados do Bóson de Higgs	113
Tabela 75 – Hiperparâmetros mais Relevantes para a Métrica no ResNet	114
Tabela 76 – Hiperparâmetros mais Relevantes para o Tempo no ResNet	114

Tabela 77 – Hiperparâmetros mais Relevantes para a Métrica no TabNet	115
Tabela 78 – Hiperparâmetros mais Relevantes para o Tempo no TabNet	115
Tabela 79 – Hiperparâmetros mais Relevantes para a Métrica no TabTransformer	115
Tabela 80 – Hiperparâmetros mais Relevantes para o Tempo no TabTransformer	115
Tabela 81 – Hiperparâmetros mais Relevantes para a Métrica no FT-Transformer	116
Tabela 82 – Hiperparâmetros mais Relevantes para o Tempo no FT-Transformer	116
Tabela 83 – Hiperparâmetros mais Relevantes para a Métrica no SAINT	116
Tabela 84 – Hiperparâmetros mais Relevantes para o Tempo no SAINT	117
Tabela 85 – Hiperparâmetros mais Relevantes para a Métrica no FastFormer	117
Tabela 86 – Hiperparâmetros mais Relevantes para o Tempo no FastFormer	117
Tabela 87 – Otimizador Escolhido na Melhor Combinação de Hiperparâmetros	118

LISTA DE ABREVIATURAS E SIGLAS

ANOVA	<i>Analysis of Variance</i>
Ativ.	Ativação
AUPR	<i>Area Under the Precision-Recall Curve</i>
AUROC	<i>Area Under the Receiver Operating Characteristics</i>
CERN	European Organization for Nuclear Research
CPU	<i>Central Processing Unit</i>
DM-SNR	<i>Dispersion-Measure-Signal-to-Noise-Ratio</i>
Embed.	<i>Embedding</i>
fANOVA	<i>functional ANOVA</i>
FF	<i>Feed-Forward</i>
FT-Transformer	<i>Feature Tokenizer Transformer</i>
GeV	giga Elétron-Volt
GLU	<i>Gated Linear Unit</i>
GPU	<i>Graphic Processing Unit</i>
KNN	<i>K Nearest Neighbors</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LightGBM	<i>Light Gradient Boosting Machine</i>
LS	<i>Least Squares</i>
MCC	Coeficiente de Correlação de Matthews
MLogP	<i>Moriguchi Octanol-Water Partition Coefficient</i>
MLP	<i>Multilayer Perceptron</i>
ON1V	<i>Overall Modified Zagreb Index of Order 1 by Valence Vertex Degrees</i>
PCD	<i>Difference Between Multiple Path Count and Path Count</i>
piPC09	<i>Molecular Multiple Path Count of Order 9</i>
ResNet	<i>Residual Network</i>
SAINT	<i>Self-Attention and Intersample Attention Transformer</i>
SGD	<i>Stochastic Gradient Descent</i>

SHAP	<i>Shapley Additive Explanations</i>
SMS	<i>Short Message Service</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i>
TLU	<i>Threshold Logic Unit</i>
XGBoost	<i>Extreme Gradient Boosting</i>

LISTA DE SÍMBOLOS

μ	Média
σ	Desvio Padrão
τ	Tau
C	Carbono
g	grama
kg	quilograma
L	litro
m ³	metro cúbico
MPa	Megapascal
N	Nitrogênio
O	Oxigênio
s	segundo
SO ₂	Dióxido de Enxofre

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivos	18
1.2	Modelos	19
1.2.1	Regressões Penalizadas	19
1.2.2	K Vizinhos Mais Próximos	21
1.2.3	Árvores de Decisão	21
1.2.4	Ensembles, Bagging e Boosting	22
1.2.5	Máquina de Vetores de Suporte	24
1.2.6	Perceptron de Múltiplas Camadas	26
1.2.7	Residual Networks	28
1.2.8	TabNet	29
1.2.9	Redes Neurais Profundas Baseadas em Transformers	30
1.2.9.1	TabTransformer	34
1.2.9.2	FT-Transformer	35
1.2.9.3	FastFormer	36
1.2.9.4	SAINT	37
1.3	Interpretabilidade	39
1.3.1	Interpretabilidade para Redes Neurais	40
1.4	Teste de Hipótese para Classificadores Binários	41
2	BANCOS DE DADOS	43
2.1	Separação dos Dados	46
2.2	Tratamento dos Dados	47
3	ESTUDOS	49
3.1	Seleção dos Hiperparâmetros	50
4	EXPERIMENTOS	51
4.1	Atrasos no Sistema Único de Saúde	51
4.2	Preço de Diamantes	56
4.3	Massa Invariante na Colisão de Muóns	61
4.4	Temperatura Crítica de Supercondutores	66
4.5	Identificação de Pulsares	72
4.6	Identificação de Espécies de Feijões	76
4.7	Telemarketing Banco	80
4.8	Notas para Vinhos	84
4.9	Anéis de Abalones	89
4.10	Cardiotocografia	94
4.11	Resistência à Compressão do Concreto	98
4.12	Fator de Bioconcentração	103
4.13	Deteccção do Bóson de Higgs	107
5	RESULTADOS TESTES DE HIPÓTESE	111

6	IMPORTÂNCIA DOS HIPERPARÂMETROS	114
7	DISCUSSÃO E TRABALHOS FUTUROS	119
7.1	Trabalhos Futuros	119
	REFERÊNCIAS	121

1 INTRODUÇÃO

Dados tabulares são a forma mais comum de apresentação de informações em problemas reais (CHUI et al., 2018). Muitas indústrias críticas, como a de serviços financeiros, saúde e logística tem uma forte dependência de dados estruturados no formato de tabelas (SOMEPALI et al., 2021).

Algumas abordagens já estabelecidas na literatura para o estudo dessa classe de dados são as Regressões, que incluem os Modelos Lineares Generalizados, com ou sem penalizações, e os Modelos Aditivos Generalizados; Máquina de Vetores de Suporte, Árvores de Decisão, seus *ensembles* como o *Bagging* e *Gradient Boosted Trees*, além de técnicas que exploram as dependências específicas para o conjunto de dados em questão, como as de Séries Temporais e de Estatística Espacial.

Dados tabulares são únicos de várias formas que os impediram de usufruir do avanço massivo das técnicas de Aprendizado Profundo em problemas de visão e linguagem. Primeiramente, eles contêm variáveis heterogêneas, compostas por uma mistura de dados contínuos, categóricos e ordinais, sendo que esses valores podem ser independentes ou correlacionados. Secundariamente, não há uma posição inerente a informação apresentada, ou seja, a ordem das colunas é arbitrária, diferindo dos textos em que os *tokens* são discretos e a sua ordem impacta o sentido semântico. Também difere de imagens onde *pixels* são comumente contínuos e são correlacionados por proximidade. (SOMEPALI et al., 2021)

Recentemente surgiram propostas com resultados promissores que visam aplicar Redes Neurais Profundas nessa categoria de dados estruturados, delas se destacam o TabNet (ARIK; PFISTER, 2020), TabTransformer (HUANG et al., 2020), FT-Transformer (GORISHNIY et al., 2021) e SAINT (SOMEPALI et al., 2021).

Elas fazem uso de avanços da área, como *Transformers*, que descobrem relações entre as variáveis, e Atenção Sequencial, que realiza uma seleção para as variáveis mais relevantes em cada parte do treinamento do modelo (como aplicado no TabNet), permitindo inclusive elencar as variáveis mais importantes auxiliando na interpretabilidade do resultado final, e também das amostras a serem consideradas em cada época de treino (demonstrado no SAINT).

1.1 OBJETIVOS

Será comparado, em conjuntos de dados reais, o desempenho dos modelos de Regressão Penalizada, K Vizinhos Mais Próximos (FIX; HODGES, 1951), Árvore de Decisão, *Random Forest* (HO, 1995), *Extra Trees* (GEURTS; ERNST; WEHENKEL, 2006), XGBoost (CHEN; GUESTRIN, 2016), LightGBM (KE et al., 2017), CatBoost (PROKHORENKOVA et al., 2019), Máquina de Vetores de Suporte (CORTES; VAPNIK, 1995), *Multilayer Perceptron*, ResNet (HE et al., 2016), TabNet (ARIK; PFISTER, 2020), TabTransformer (HUANG et al., 2020), FT-Transformer (GORISHNIY et al., 2021), FastFormer (WU et al., 2021) e SAINT (SOMEPELLI et al., 2021).

Além do desempenho em métricas relacionadas à qualidade da predição, também será analisada a performance no tempo, que irá mensurar o tempo de treinamento de cada modelo com o melhor conjunto de hiperparâmetros escolhidos. Por último, também se deseja mensurar a importância das variáveis em cada modelo, comparando a forma com que esses métodos tão distintos têm de entender os problemas em questão.

1.2 MODELOS

Em um cenário típico, temos uma variável resposta medida, usualmente quantitativa (como os preços de ações negociadas na bolsa de valores) ou qualitativa (como se o paciente tem ou não um ataque cardíaco) que queremos prever baseado em um conjunto de variáveis explicativas (como dieta e medidas clínicas). Temos um conjunto de dados para treino, no qual observamos os resultados e medidas das variáveis explicativas para um grupo de amostras (pessoas, por exemplo). Usando esses dados podemos construir um modelo de predição que nos permitirá prever o resultado de amostras ainda não observadas. (HASTIE; TIBSHIRANI; FRIEDMAN, 2001)

Os problemas em estudos se encaixam na categoria de aprendizado supervisionado, pois temos as informações verdadeiras a cerca da variável resposta de cada amostra e podemos então comparar a predição com a realidade.

De forma geral, podemos representar todos os modelos como uma função das variáveis explicativas que retornam um único valor como a imagem dessa função:

$$f(X_i) = \hat{y}_i \quad (1)$$

Sendo X_i a matriz composta pelos valores das variáveis explicativas consideradas na i -ésima amostra e \hat{y}_i o valor predito para ela.

Nenhum método supera todos os outros em todos os possíveis conjuntos de dados. Logo é uma tarefa importante decidir para um conjunto de dados em questão qual método produz os melhores resultados. (JAMES et al., 2013)

1.2.1 REGRESSÕES PENALIZADAS

Dentre os métodos de penalização para regressões se destacam o *Ridge* (HOERL; KENNARD, 1970), LASSO (TIBSHIRANI, 1994), e sua combinação *Elastic Net* (ZOU; HASTIE, 2005), funcionando de forma semelhante, somando à função de perda escolhida (como o erro quadrático médio para problemas de regressão) uma outra função de um parâmetro positivo e dos coeficientes da regressão. Minimizando essa nova função de perda se obtém um ajuste que se adequa aos dados, mas que também possua um menor número de parâmetros ou um menor coeficiente para eles.

A penalização, também chamada de regularização, tem como objetivo aumentar a capacidade de generalização dos modelos, forçando a escolha de quais parâmetros devem ter maior peso, evitando assim o sobreajuste.

A penalização *Ridge* soma à função de perda um parâmetro positivo λ multiplicado pelo somatório do quadrado de cada coeficiente. Como se deseja minimizar essa nova função, quanto maior for o parâmetro escolhido, menores serão os coeficientes. A penalização *Ridge* que será somada a função de perda original é definida como:

$$\lambda \sum_{i=1}^p \beta_i^2 \quad (2)$$

Sendo β_i o coeficiente atribuído para a i -ésima variável explicativa e p o número total de variáveis independentes no modelo.

Zou e Hastie (2005) descrevem seu funcionamento como a minimização dos resíduos da soma dos quadrados sujeitados a uma limitação na norma L_2 dos coeficientes, obtendo uma melhor predição através do *trade-off* entre viés e variância, mas que não consegue selecionar variáveis, pois nenhum coeficiente irá ser reduzido para zero.

Essa limitação se dá devido a que qualquer número real, diferente de zero, que pertença ao intervalo aberto $(-1, 1)$ elevado ao quadrado se tornará menor, em módulo, do que seu valor original; logo, para valores pequenos dos coeficientes a penalização Ridge focaria em reduzir outros coeficientes maiores, não podendo realizar uma seleção de variáveis.

O LASSO é construído de forma semelhante ao *Ridge* se diferindo unicamente por substituir o somatório do quadrado dos coeficientes pelo somatório de seus valores absolutos. A penalização LASSO é definida como:

$$\lambda \sum_{i=1}^p |\beta_i| \quad (3)$$

A vantagem do LASSO é que, por contar com o absoluto dos coeficientes pode reduzi-los para zero, podendo assim realizar um processo de seleção de variáveis.

Elastic Net é a combinação convexa das penalizações introduzidas pelo LASSO e *Ridge*, com o objetivo de obter tanto um processo que possa selecionar variáveis, com um melhor equilíbrio entre o viés e variância. A penalização *Elastic Net* é definida como:

$$\alpha \lambda \sum_{i=1}^p \beta_i^2 + (1 - \alpha) \lambda \sum_{i=1}^p |\beta_i| \quad (4)$$

Sendo α um parâmetro que pertence ao intervalo $[0,1]$ e determina a proporção das penalizações *Ridge* e LASSO.

Outra abordagem foi proposta por Hastie, Tibshirani e Tibshirani (2017), sendo uma variação do *Relaxed Lasso* de Meinshausen (2007), onde em um problema de regressão se ajustam parâmetros λ e γ , o primeiro é não-negativo e controla a seleção de variáveis de forma idêntica ao LASSO, já o segundo pertence ao intervalo $[0,1]$ e multiplica os coeficientes inicialmente reduzidos por λ podendo reduzir (caso seja diferente de 1, situação onde teríamos o LASSO convencional) sua penalização caso melhore o ajuste. Os novos coeficientes podem ser descritos pela equação:

$$\beta^{relaxed}(\lambda, \gamma) = \gamma \beta^{LASSO}(\lambda) + (1 - \gamma) \beta^{LS}(\lambda) \quad (5)$$

Sendo $\beta^{relaxed}$ o vetor que contém os coeficientes finais, β^{LASSO} o vetor dos coeficientes selecionados pelo LASSO e β^{LS} o vetor dos coeficientes que obteríamos em uma regressão pelo Método dos Mínimos Quadrados utilizando somente as variáveis selecionadas pelo LASSO.

Nos experimentos conduzidos por Hastie, Tibshirani e Tibshirani (2017) o método do *Relaxed Lasso* demonstrou consistentemente uma melhor performance, do que outras regressões penalizadas, nos ajustes conduzidos em conjuntos de dados simulados de diferentes dimensões e tamanho amostral.

1.2.2 K VIZINHOS MAIS PRÓXIMOS

K Vizinhos Mais Próximos (FIX; HODGES, 1951) é uma técnica não-paramétrica onde se escolhe um número natural, denotado por K , criando um preditor para um ponto qualquer através de uma função do valor assumido pela variável resposta dos K pontos vizinhos mais próximos a ele. Esses vizinhos são escolhidos através da proximidade das variáveis explicativas desse ponto específico e dos dados que fazem parte do conjunto de treino.

Suas variantes se dão pela escolha da métrica de distância (Euclidiana, Minkowski, *Manhattan*, entre outras), na função que vai ser aplicada aos vizinhos (escolhas mais usuais são a moda, a média aritmética e a média ponderada pelo inverso da distância) e do algoritmo usado para mensurar os pontos.

Quando K é igual a um, a fronteira de decisão é exageradamente flexível e acha padrões nos dados que não correspondem à fronteira de decisão de um classificador Bayesiano. Isso corresponde a um classificador de baixo viés e alta variância. Conforme K cresce o método se torna menos flexível e produz uma fronteira de decisão que se aproxima da linearidade. Isso corresponde a um classificador de baixa variância, mas alto viés. (JAMES et al., 2013)

1.2.3 ÁRVORES DE DECISÃO

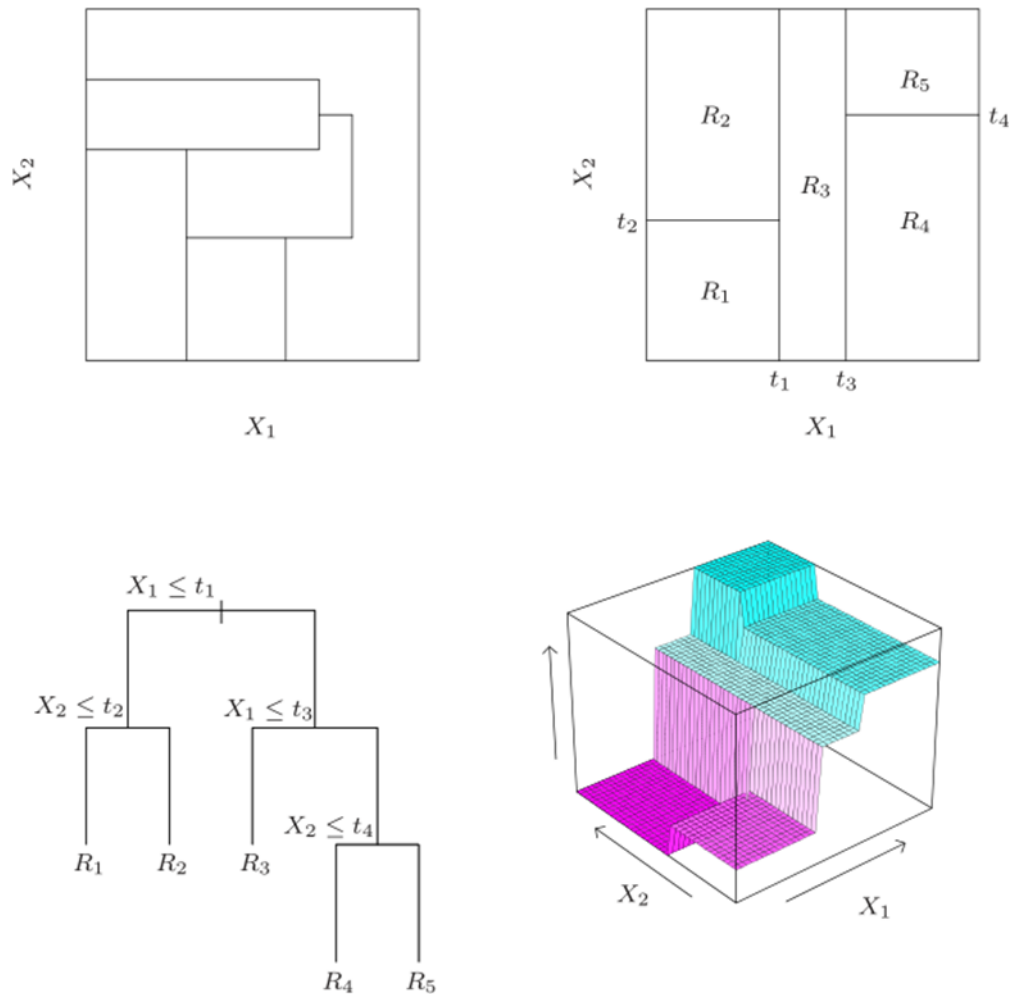
Métodos baseados em árvores particionam o espaço das variáveis em uma série de retângulos e ajustam um modelo simples (como uma constante) em cada um deles. Eles são conceitualmente simples, mas ainda assim poderosos. (HASTIE; TIBSHIRANI; FRIEDMAN, 2001)

Uma Árvore de Decisão (*Decision Tree*) é composta por níveis, denominados por “nós” ou “raízes”, em que variáveis explicativas são escolhidas e neles se determina um ponto de corte que separa as amostras em diferentes grupos, cada separação de um nó é conhecida como “folha”. Após todas as separações serem realizadas é então tomada uma função dos elementos que pertencem a cada grupo.

Nela a separação de cada nó pode ser feita através da maximização de um critério de pureza, como Gini ou Entropia Cruzada (para problemas de classificação) ou o Erro Quadrático Médio (para regressões), visando encontrar a melhor separação entre os grupos de amostras.

Utilizando um número suficientemente grande de nós é possível atingir uma classificação perfeita ou próxima disso para qualquer número de amostras do conjunto de treino, mas essa escolha leva a uma perda da capacidade de generalização do modelo. Uma forma de contornar essa limitação é através da formação de árvores menores; comparando métricas de predição em um conjunto de validação, ou seja, amostras em que o modelo não foi treinado, para diferentes quantidades de raízes; outra forma é através do uso de *ensembles*.

Figura 1 – Representação de Árvores de Decisão



Fonte: Hastie, Tibshirani e Friedman (2001)

A figura 1 demonstra a formulação de uma árvore de decisão com o particionamento do espaço de variáveis em múltiplas regiões.

1.2.4 ENSEMBLES, BAGGING E BOOSTING

A seguinte analogia proposta por Géron (2017, p. 193) fornece intuição sobre o funcionamento dos *ensembles*: Suponha que você tenha uma moeda levemente enviesada que possui 51% de chance de ter o resultado Cara e 49% para Coroa. Se ela for lançada 1000 vezes, é esperado que se obtivesse 510 Caras e 490 Coroas, portanto, uma maioria de Caras. Calculando podemos concluir que a probabilidade de termos uma maioria de Caras nesses lançamentos é de aproximadamente 75%. Quanto maior for o número de lançamentos, maior será a probabilidade de que a maioria seja de Caras (conforme o número de lançamentos tende a infinito é esperado que essa probabilidade convirja para 1, segundo a Lei dos Grandes Números).

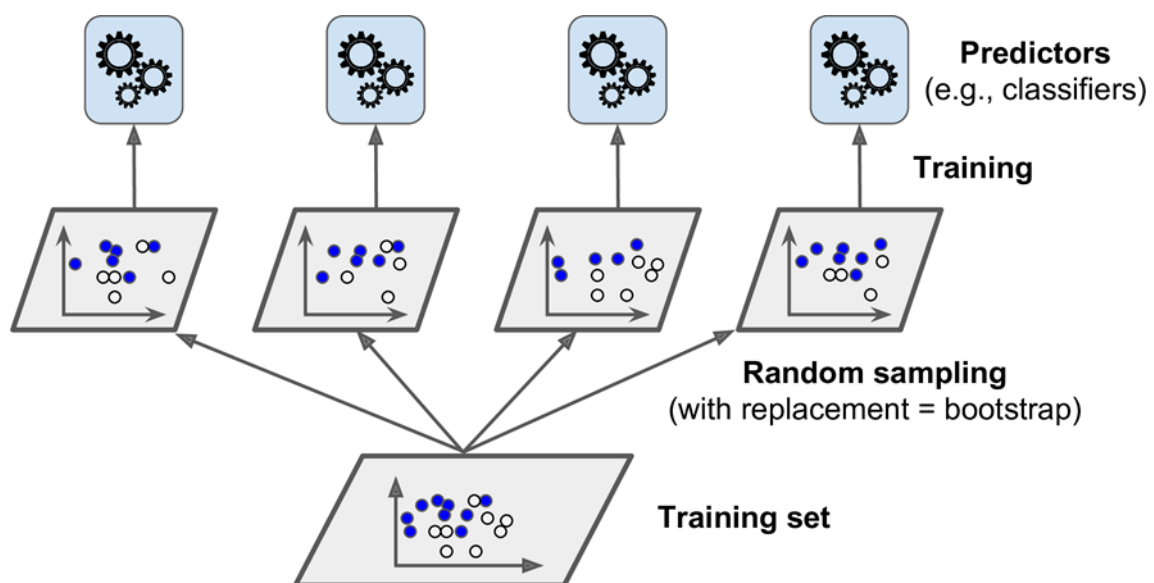
De forma similar, suponha que seja construído um *ensemble* com 1000 classificadores que estão individualmente corretos 51% do tempo, se predizermos a categoria com a maior parte dos votos estaríamos corretos 75% das vezes. Entretanto, isso somente é verdade se todos os classificadores são independentes,

fazendo erros menos correlacionados, o que claramente não é o caso dado que são treinados nos mesmos dados. (GÉRON, 2017)

O desempenho de um *ensemble* é função da performance preditiva de cada preditor individualmente, das correlações entre eles e de quantos deles foram desenvolvidos.

Random Forest é um *ensemble* de árvores de decisão, formado por um processo conhecido como *Bagging* (termo amálgama de *Bootstrap Agreggating*), onde paralelamente são calculados diversos preditores, na forma de árvores de decisão, e então realizada uma função, geralmente a média (para problemas de regressão) ou a moda (para problemas de classificação), sobre o resultado delas.

Figura 2 – Representação do Processo de Bagging



Fonte: Géron (2017)

A figura 2 demonstra o processo do *bagging*, em que são realizadas amostragens com reposição dos dados originais de treino e que serão processados separadamente pelas árvores de decisão.

A função do *bootstrap* é evitar que cada classificador seja treinado exatamente com os mesmos dados, gerando assim uma menor correlação entre eles. Outra forma comum de criar esse efeito é de limitar e atribuir aleatoriamente o número de variáveis explicativas que cada um deles possa usar.

A adição de elementos de aleatoriedade diminui a correlação entre cada preditor. Outra forma de alcançar isso foi proposta no algoritmo *Extra Trees*, muito semelhante ao *Random Forest*, mas nele ponto de separação entre as folhas é definido aleatoriamente.

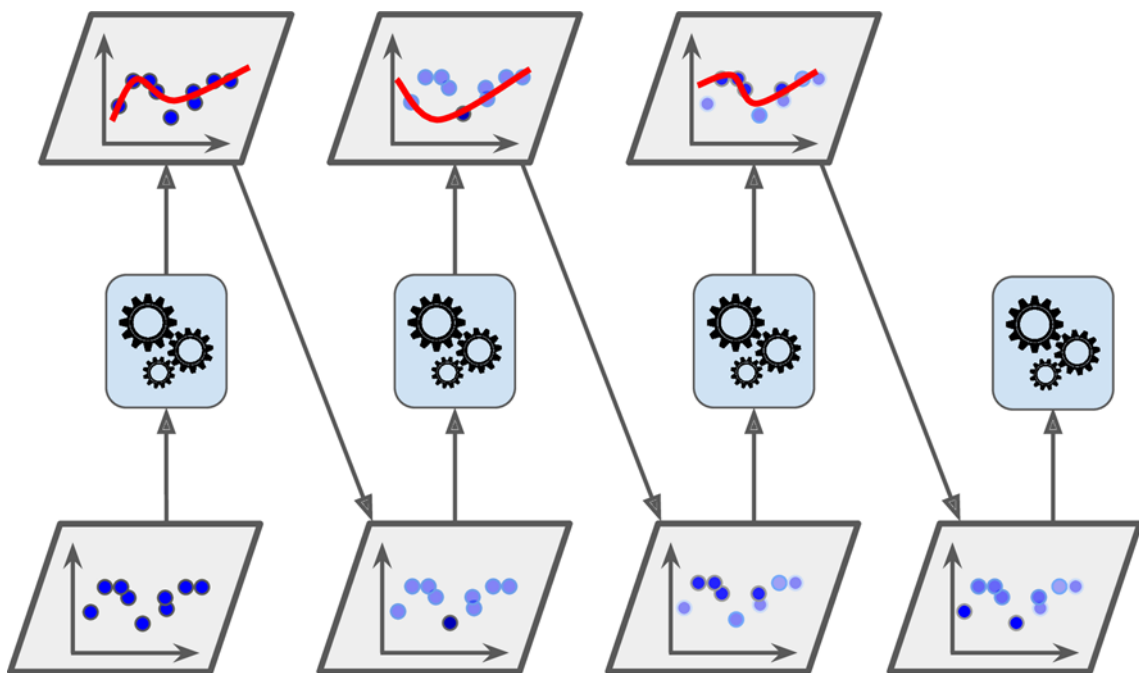
Geurts, Ernst e Wehenkel (2006) justificam essa escolha no *Extra Trees*, devido a que no *trade-off* entre viés e variância, os pontos de corte aleatórios reduzem mais severamente a variância do que no *Random Forest*, aumentando a performance do algoritmo para variáveis de extremo ruído, além da diminuição do custo computacional.

XGBoost, LightGBM, CatBoost são variações populares das técnicas de *Gradient Boosted Trees*, onde se criam árvores de decisão, geralmente de pequena profundidade, e se aplica sobre elas um processo iterativo que gera um preditor melhor do que cada uma separadamente,

dessa forma também é um *ensemble*, mas se difere do *Bagging* realizando o processo de formação das árvores de forma sequencial.

Prokhorenkova et al. (2019) caracterizam o processo de *Gradient Boosting* como a criação iterativa de uma sequência de aproximações $F^t : \mathbb{R}^m \rightarrow \mathbb{R}$, com $t = \{1, 2, \dots\}$; seguindo um algoritmo ganancioso. Então F^t é obtido pela aproximação anterior de F^{t-1} de uma forma aditiva: $F^t = F^{t-1} + \alpha h^t$, onde α é o tamanho do passo em cada etapa do algoritmo, h^t é uma função $\mathbb{R}^m \rightarrow \mathbb{R}$ de uma família de funções H que visa reduzir o valor esperado da função de perda.

Figura 3 – Representação do Processo de Boosting



Fonte: Géron (2017)

A figura 3 ilustra o processo de *boosting* com árvores de decisão sendo treinadas com modificações dos bancos de dados utilizados nas árvores anteriores.

Dado o modelo presente, uma nova árvore é ajustada sobre seus resíduos. E então adicionamos essa nova árvore de decisão a função de ajuste e atualizamos seus resíduos. (JAMES et al., 2013)

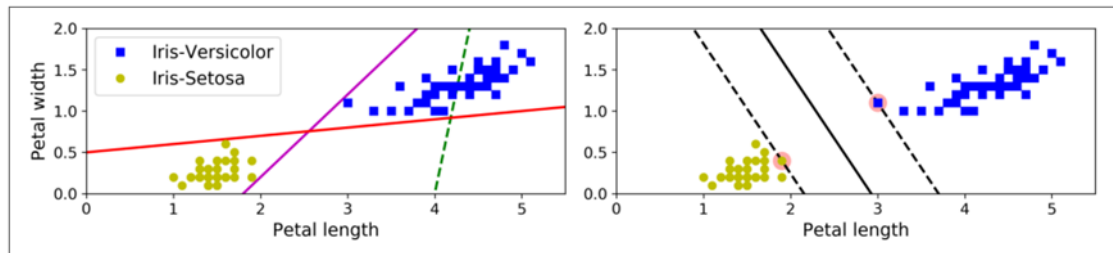
1.2.5 MÁQUINA DE VETORES DE SUPORTE

Amostras que pertencem a duas classes não são necessariamente separáveis por um hiperplano. De fato, mesmo que exista esse hiperplano que separa perfeitamente, existem situações onde um classificador baseado nele pode não ser desejável. Esse modelo classificaria perfeitamente as observações de treino, o que levaria a uma alta sensibilidade a cada uma delas. (JAMES et al., 2013)

Esse hiperplano de margem máximo não seria satisfatório – primeiramente por possuir uma pequena margem. A distância de uma observação ao hiperplano pode ser vista como uma medida da confiança que ela foi classificada corretamente. Além disso, essa alta sensibilidade a cada observação pode levar ao sobreajuste. (JAMES et al., 2013)

Nesse caso, estaríamos dispostos a considerar um classificador baseado em um hiperplano que não separa corretamente cada amostra, com o intuito que possa ser válido errar em poucas delas, para acertar em todas as restantes. O Classificador de Vetores de Suporte, muitas vezes denominado como “*Soft Margin Classifier*”, faz exatamente isso. (JAMES et al., 2013)

Figura 4 – Representação das possíveis separações de classes pelo SVM



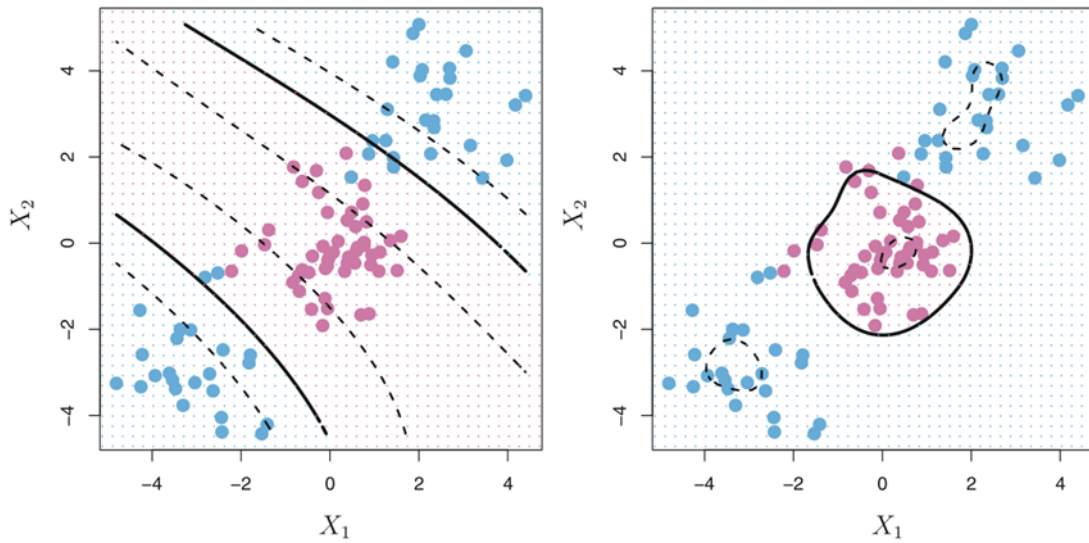
Fonte: Géron (2017)

A figura 4 demonstra possíveis hiperplanos para a classificação de amostras em grupos distintos, com a seleção final de um hiperplano que maximiza a distância da amostra de cada classe mais próxima, esses pontos são assim denominados como o suporte da fronteira de decisão.

O Classificador de Vetores de Suporte é uma abordagem natural para uma classificação binária, se a fronteira entre as classes for linear. Entretanto, na prática é comum enfrentar fronteiras de classe não-lineares. Fica claro que nelas o Classificador de Vetores de Suporte ou qualquer classificador linear teria uma performance desfavorável. (JAMES et al., 2013)

A Máquina de Vetores de Suporte é uma extensão do Classificador de Vetores de Suporte, através do crescimento do espaço de variáveis através do uso de *kernels*. (...) Podemos querer aumentar o espaço de variáveis para acomodar uma separação não-linear entre as classes, o uso de *kernels* é somente uma forma computacionalmente eficiente para isso. (JAMES et al., 2013)

Figura 5 – SVM com um Kernel Polinomial de Grau 3 (esquerda). Uso de um Kernel Radial (direita)



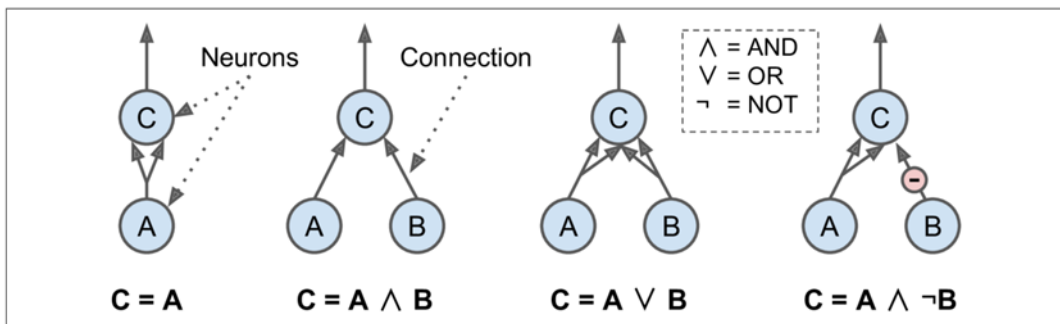
Fonte: James et al. (2013)

A figura 5 demonstra o uso de *kernels* não-lineares para a delimitação da fronteira de decisão em um modelo de classificação binária.

1.2.6 PERCEPTRON DE MÚLTIPLAS CAMADAS

McCulloch e Pitts (1943) propuseram um modelo muito simples de um neurônio biológico, que ficou conhecido posteriormente como neurônio artificial: Ele possui uma ou mais entradas binárias e uma única saída binária. O neurônio artificial simplesmente ativa sua saída quando mais do que um dado número de suas entradas são ativadas. McCulloch e Pitts mostraram que mesmo esse modelo simplificado é capaz de construir uma rede de neurônios artificiais que poderia computar qualquer proposição lógica. (GÉRON, 2017)

Figura 6 – Neurônios Artificiais Computando Operadores Lógicos

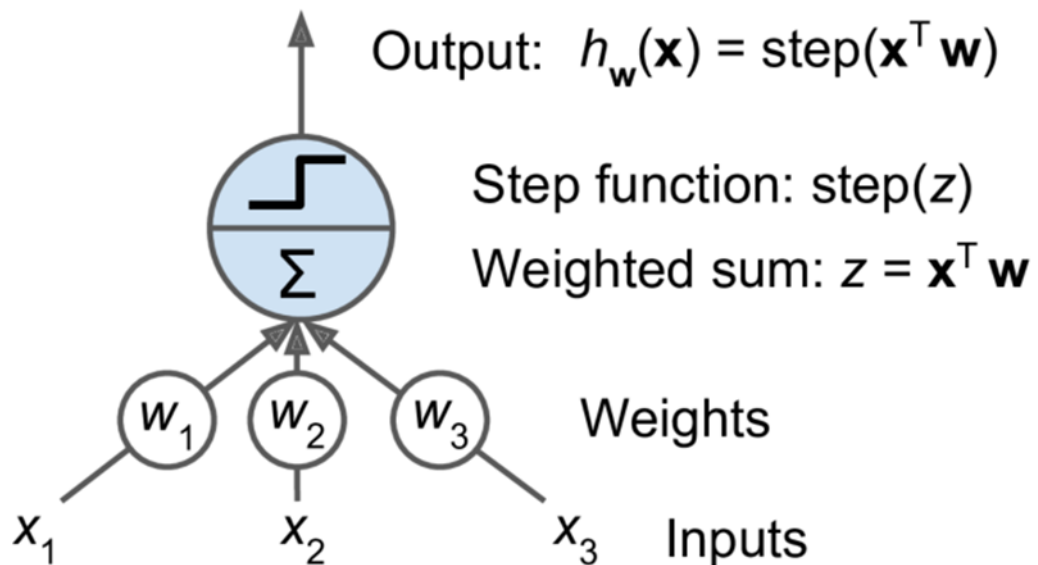


Fonte: Géron (2017)

A figura 6 ilustra esse modelo de neurônio artificial, demarcado pela letra C, com suas entradas A e B que poderão ter seus valores transformados por operações de lógica booleana.

O *Perceptron* é uma das mais simples arquiteturas de Redes Neurais Artificiais, inventado em 1957 por Frank Rosenblatt (1957). Ele é baseado em um neurônio artificial levemente diferente chamado de *Threshold Logic Unit*, ou também *Linear Threshold Unit*: As entradas são agora números (em vez do anterior binário) e cada uma é associada a um peso. O TLU computa uma soma ponderada pelos pesos das entradas e então aplica uma função degrau nela. (GÉRON, 2017)

Figura 7 – Modelo do TLU



Fonte: Géron (2017)

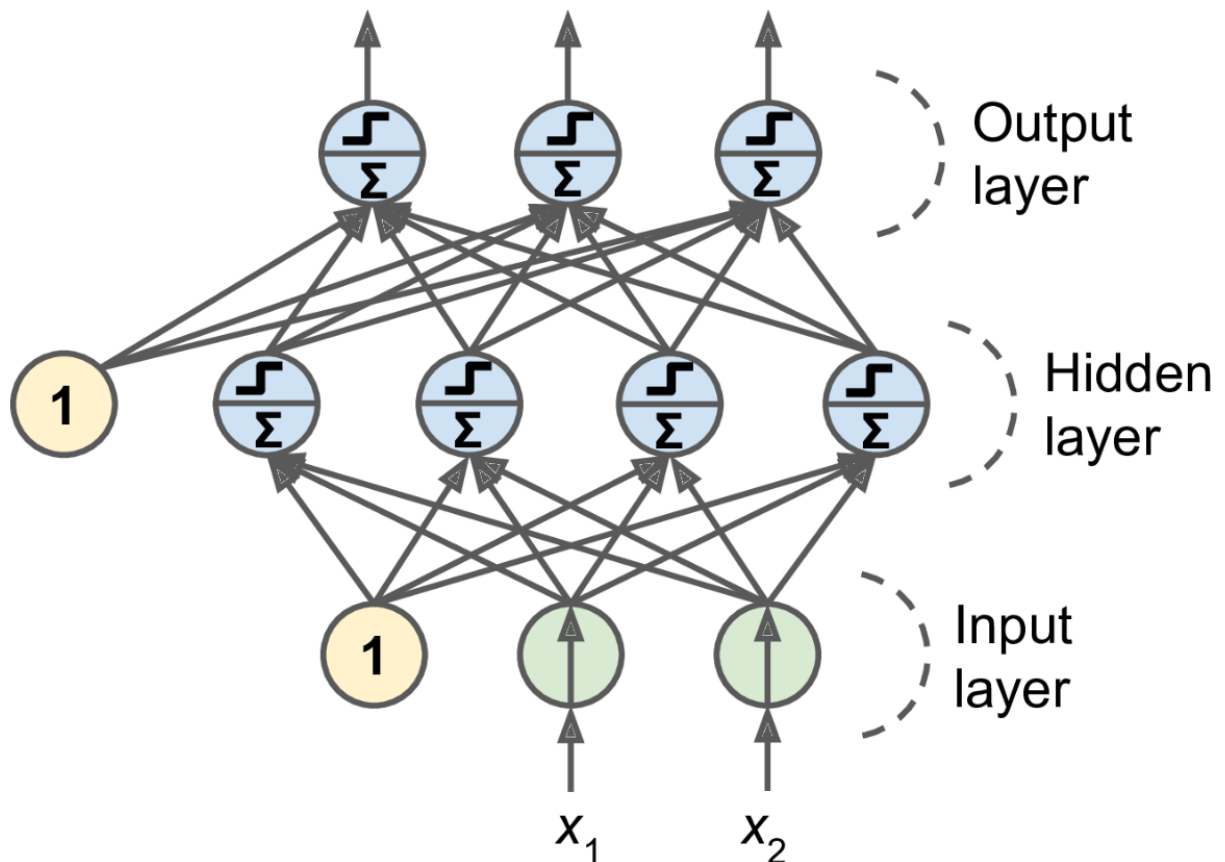
A figura 7 ilustra o modelo de um TLU, com suas entradas denotadas pela letra X , os pesos associados a elas por W , e por fim sua saída como resultado da função de ativação aplicada à combinação linear das entradas e pesos.

Um único TLU pode ser usado para uma classificação linear binária. Ele computa uma combinação linear das entradas, se o resultado supera um dado limite, ele escolhe a classe positiva, caso contrário escolhe a negativa (de forma similar a uma Regressão Logística ou SVM linear). Treinar o TLU nesse caso significa encontrar os melhores valores para os pesos. (GÉRON, 2017)

Um *Perceptron* é composto de uma única camada de TLUs, cada um deles conectado com todas as entradas. Quando todos os neurônios de uma camada estão conectados com todos da camada anterior, isso é chamado de camada totalmente conectada ou camada densa. (GÉRON, 2017)

Um Perceptron de Múltiplas Camadas é composto por uma camada de entradas, uma ou mais de TLUs, chamadas de camadas ocultas (*hidden layers*), e uma camada final de TLUs chamada de camada de saídas. (GÉRON, 2017)

Figura 8 – Representação de um Perceptron de Múltiplas Camadas com a Adição de Neurônios de Viés



Fonte: Géron (2017)

A figura 8 representa a arquitetura de um *Multilayer Perceptron* com suas camadas em sequência, cada uma composta por múltiplos *Perceptrons*.

Rumelhart, Hinton e Williams (1986) publicaram um artigo extremamente inovador introduzindo o conceito de retropropagação (*backpropagation*) para o treinamento desses algoritmos, técnica usada até os dias de hoje. Resumidamente, ela faz uso do Gradiente Descendente usando uma eficiente técnica para calcular os gradientes automaticamente: em dois passos através da rede (um avançando e um retrocedendo), o algoritmo de retropropagação é capaz de calcular os erros da rede em relação a cada parâmetro do modelo. Em outras palavras, ele encontra como cada peso e viés de cada conexão deveriam ser mudados para diminuir esse erro. Após esses gradientes serem encontrados, ele realiza o Gradiente Descendente e todo o processo recomeça até a rede convergir para a solução. (GÉRON, 2017)

O *Multilayer Perceptron* por ser o modelo mais simples de uma rede neural servirá como uma comparação para o ganho de desempenho das escolhas de design feitas nas redes neurais mais complexas.

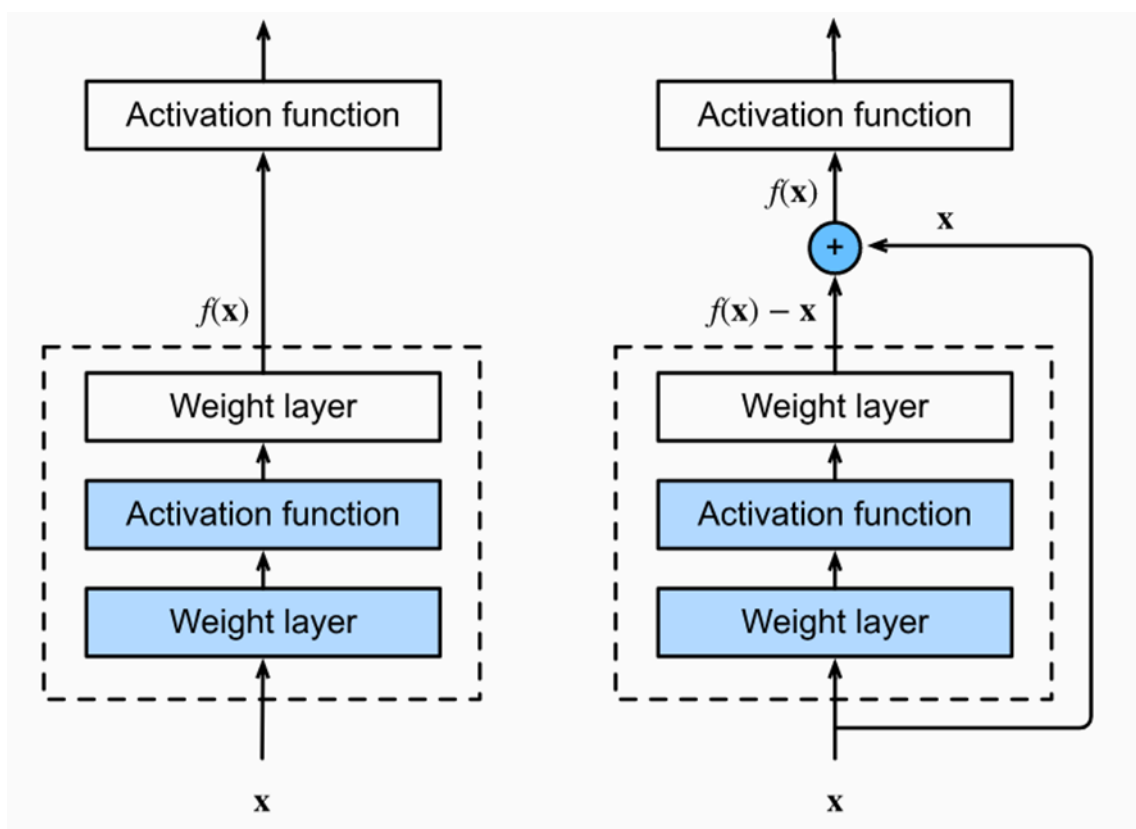
1.2.7 RESIDUAL NETWORKS

Originalmente proposto para processos de visão computacional, mas que encontrou aplicações em outros tipos de dados.

Quando redes profundas começam a convergir, um problema de degradação é exposto: conforme a profundidade da rede aumenta, a acurácia satura e degrada rapidamente. De forma inesperada, tal degradação não é causada por sobreajuste e adicionar mais camadas leva a um maior erro no treino. (HE et al., 2016)

Uma conexão residual consiste na reintrodução de representações anteriores nos passos seguintes, adicionando um tensor de saída anterior em um tensor de saída posterior, o que previne perda de informação durante o processamento dos dados. (CHOLLET, 2021, p.235) Um diagrama desse processo pode ser visto na imagem 9.

Figura 9 – Comparação da Arquitetura de uma Rede Neural Convencional (esquerda) com a de uma Residual Network (direita)



Fonte: Zhang et al. (2021)

1.2.8 TABNET

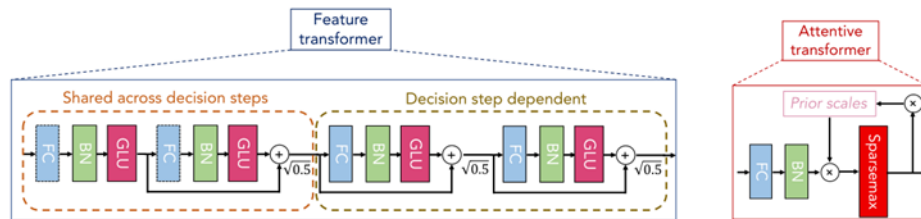
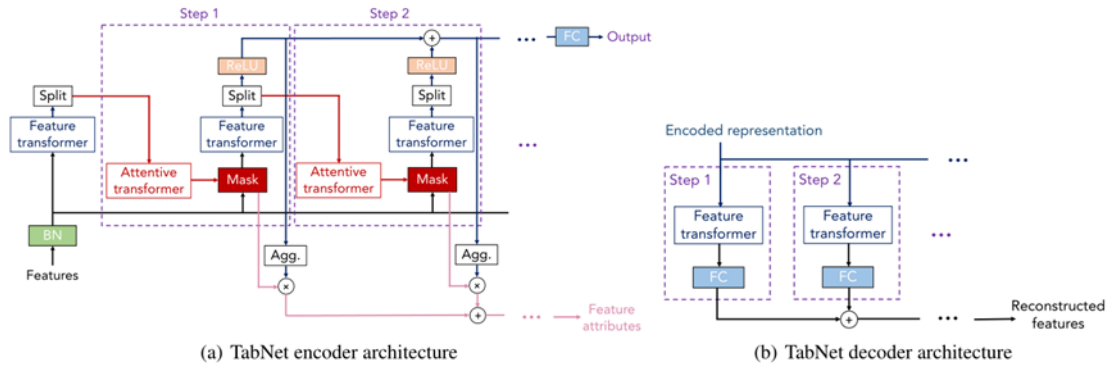
Arik e Pfister (2020) fundamentam a escolha da arquitetura do TabNet pela capacidade de partes de Redes Neurais Profundas poderem ser usadas para implementar saídas semelhantes aos espaços gerados por Árvores de Decisão. Consideram que nesse design a seleção de variáveis a serem usadas é peça-chave para obter o hiperplano da fronteira de decisão, que pode ser generalizado como a combinação linear dessas variáveis.

Esse mecanismo de Atenção tem como efeito selecionar parte dos dados e assinalar a ela uma maior relevância para uma dada etapa do treinamento, conceito similar à cognição biológica, onde seres vivos se concentram primariamente em partes de seu campo de visão para tomar decisões, como dando maior relevância aos olhos e a postura de quem está a sua frente,

ou a certos sinais sonoros, como os produzidos por predadores. Um exemplo de sua aplicação está em algoritmos de visão computacional, onde para reconhecer uma pessoa se dá mais valor ao rosto do que a outras partes do corpo humano.

Arik e Pfister (2020) descrevem a atuação do TabNet como: o uso de seleção de variáveis em cada instância do treino, processamento não-linear dessas variáveis e a representação delas através de processos que simulam *ensembles*.

Figura 10 – Arquitetura do TabNet



Fonte: Arik e Pfister (2020)

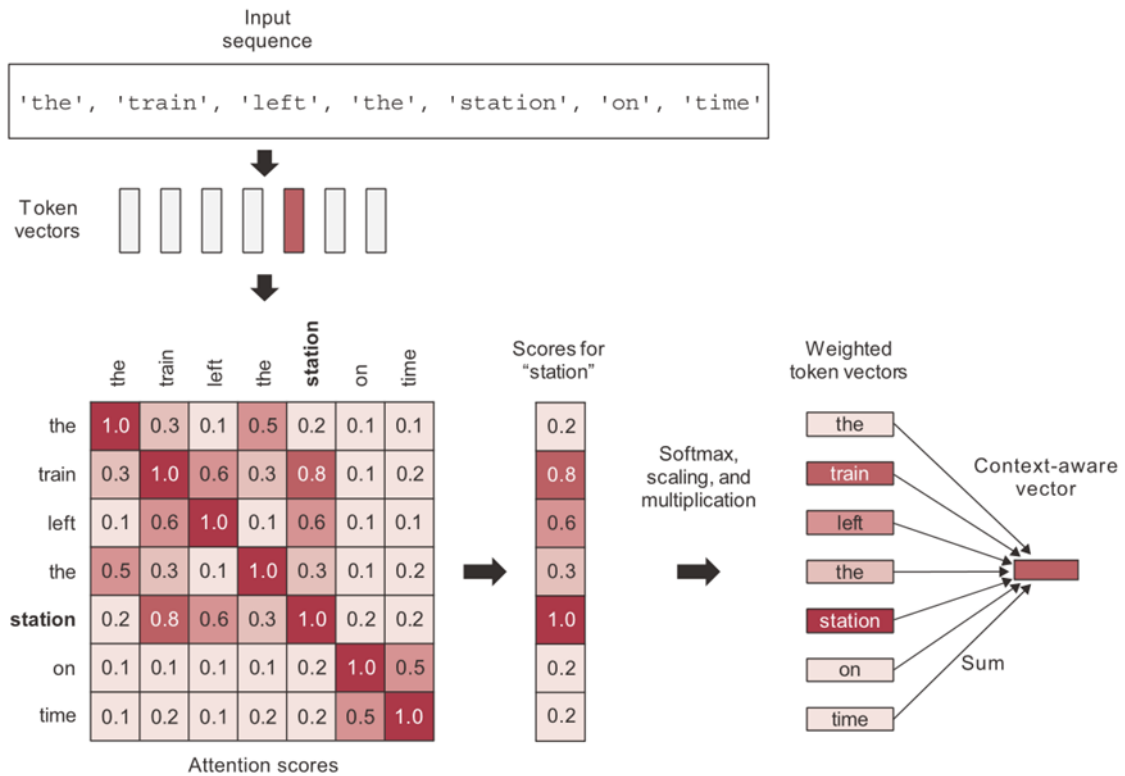
1.2.9 REDES NEURAIS PROFUNDAS BASEADAS EM TRANSFORMERS

Uma evolução da área de Aprendizado Profundo essencial para o funcionamento do TabTransformer (HUANG et al., 2020), SAINT (SOMEPALLI et al., 2021), FastFormer (WU et al., 2021) e FT-Transformer (GORISHNIY et al., 2021) é a arquitetura *Transformer* (VASWANI et al., 2017), com seu mecanismo de Atenção.

Vaswani et al. (2017) retratam o funcionamento do *self-attention*, em problemas de Processamento de Linguagem Natural, como um mecanismo relacionando diferentes posições de uma sequência de palavras com objetivo de computar uma representação dela.

O propósito do *self-attention* é modularizar a representação de um *token* através do uso das representações de *tokens* relacionados em uma sequência. Isso produz representações condizentes com seu contexto. Considere a sentença: “O trem deixou a estação pontualmente.” Agora, considere uma palavra da sentença: estação. Qual tipo de estação estamos falando? Poderia ser uma estação de rádio? Ou talvez a Estação Espacial Internacional? (CHOLLET, 2021, p.338)

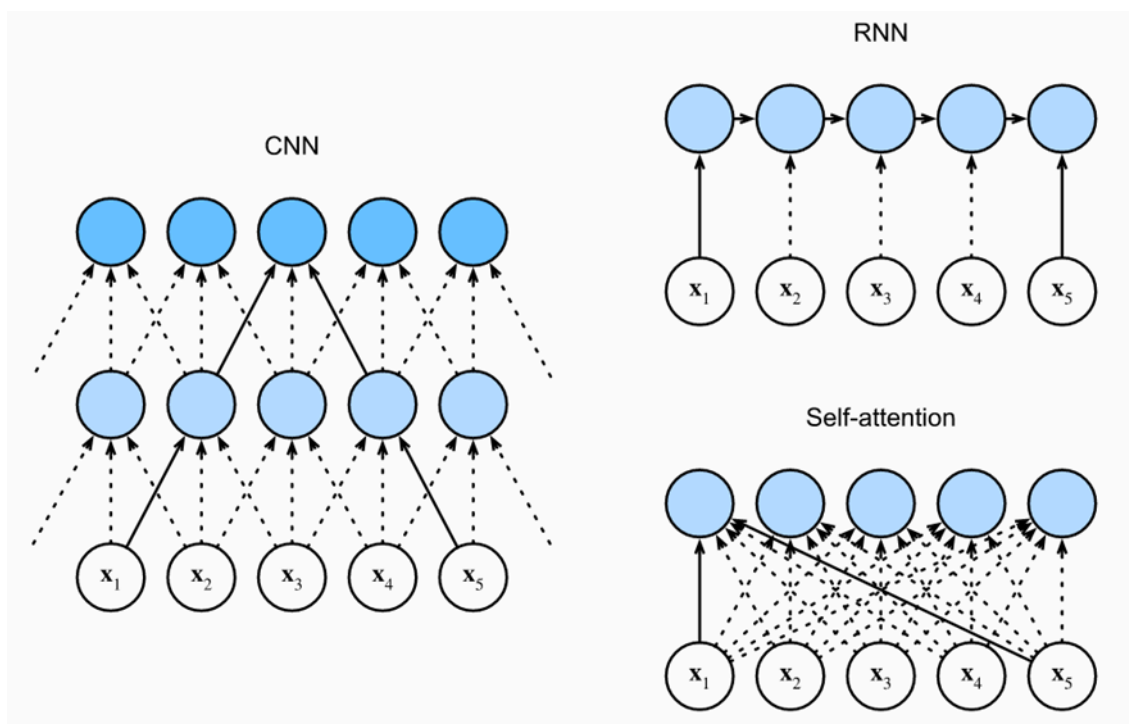
Figura 11 – Representação do Mecanismo de Self-Attention para Processamento de Linguagem Natural



Fonte: Chollet (2021)

Na imagem 11 Chollet (2021) descreve o processo de um mecanismo de Atenção, começando através de calcular a relevância entre cada palavra da sentença e o vetor da palavra “estação”, através do produto escalar dos vetores das palavras, posteriormente esses resultados são padronizados e passam por uma função *Softmax*. Após isso, é calculada a soma de todos os vetores, ponderada pelos valores das relevâncias, dando maior peso para as palavras com associação mais forte. O vetor resultante se torna então a nova representação para a palavra “estação”, incorporando o contexto que a cerca, incluindo também parte do vetor da palavra “trem” deixando claro qual tipo de estação a sentença se refere.

Figura 12 – Comparação da arquitetura de Self-Attention com a de outras redes neurais



Fonte: Zhang et al. (2021)

Pode-se perceber no diagrama da figura 12 que o mecanismo de *self-attention* alimenta cada neurônio da camada seguinte com todas as entradas da camada anterior, com o objetivo que cada um deles tenha acesso à totalidade das representações dos dados, para escolher somente um conjunto delas que será considerado mais relevante em cada neurônio da camada seguinte.

Vaswani et al. (2017) descrevem três características do mecanismo de *self-attention* que levaram seu uso na criação do *Transformer*: O total da complexidade computacional em cada camada, quanto do custo computacional pode ser paralelizado, medido como o mínimo de operações sequenciais necessárias, e por fim a distância entre as dependências de maior distância da rede, que quanto menores mais facilmente se realiza o processo de aprendizado.

O *Transformer* é composto por duas partes, um *Encoder* e um *Decoder*, a função do primeiro é capturar relações entre os dados, através da criação de *embeddings*, e a do segundo é retornar esses *embeddings* para o formato original dos dados.

Um *embedding* é qualquer função que transforme vetores de alta dimensão em um espaço de menor dimensão, geralmente escolhida com o intuito de representar alguma relação entre eles.

O *Encoder* é composto por camadas idênticas, cada uma delas com duas subcamadas. A primeira é um mecanismo de *multi-head self-attention*, e a segunda é uma *Feed-Forward Network* totalmente conectada. (VASWANI et al., 2017)

Já o *Decoder* também é composto por camadas idênticas, em adição as duas subcamadas de cada *Encoder*, ele insere uma terceira subcamada que realiza *multi-head attention* sobre as saídas do *Encoder*. (VASWANI et al., 2017)

Sobre o *multi-head attention*, Vaswani et al. (2017) descrevem seu funcionamento como a projeção linear das dimensões. Em cada uma delas é realizado o processo de Atenção parale-

lamente; então suas saídas são concatenadas e novamente projetadas. Justifica-se essa escolha como uma forma de aprender representações do subespaço gerado em diferentes posições, algo que o *Attention* convencional não conseguiria.

Uma *head* do mecanismo de *self-attention* pode ser formulada da seguinte forma:

$$Multihead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (6)$$

Sendo $Q, K, V \in \mathbb{R}^{N \times d}$ as entradas para o mecanismo de Atenção, d a dimensão interna de cada *attention head* e $W^O \in \mathbb{R}^{hd \times d}$ é uma matriz de transformação linear dos parâmetros. (WU et al., 2021)

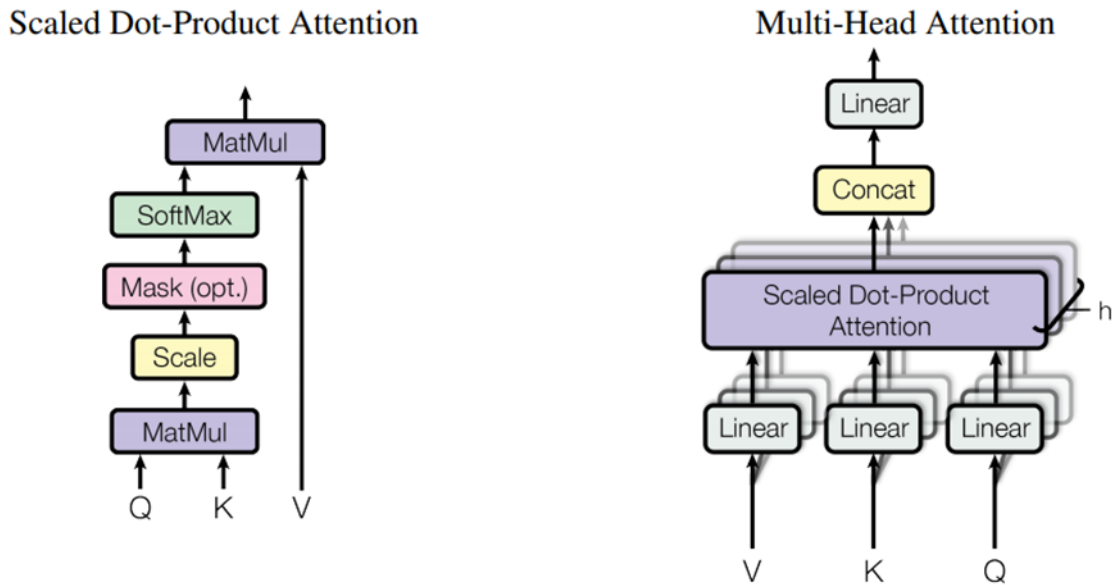
A representação aprendida por cada *attention head* pode ser formulada como:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (7)$$

$$head_i = Softmax\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d}}\right)VW_i^V \quad (8)$$

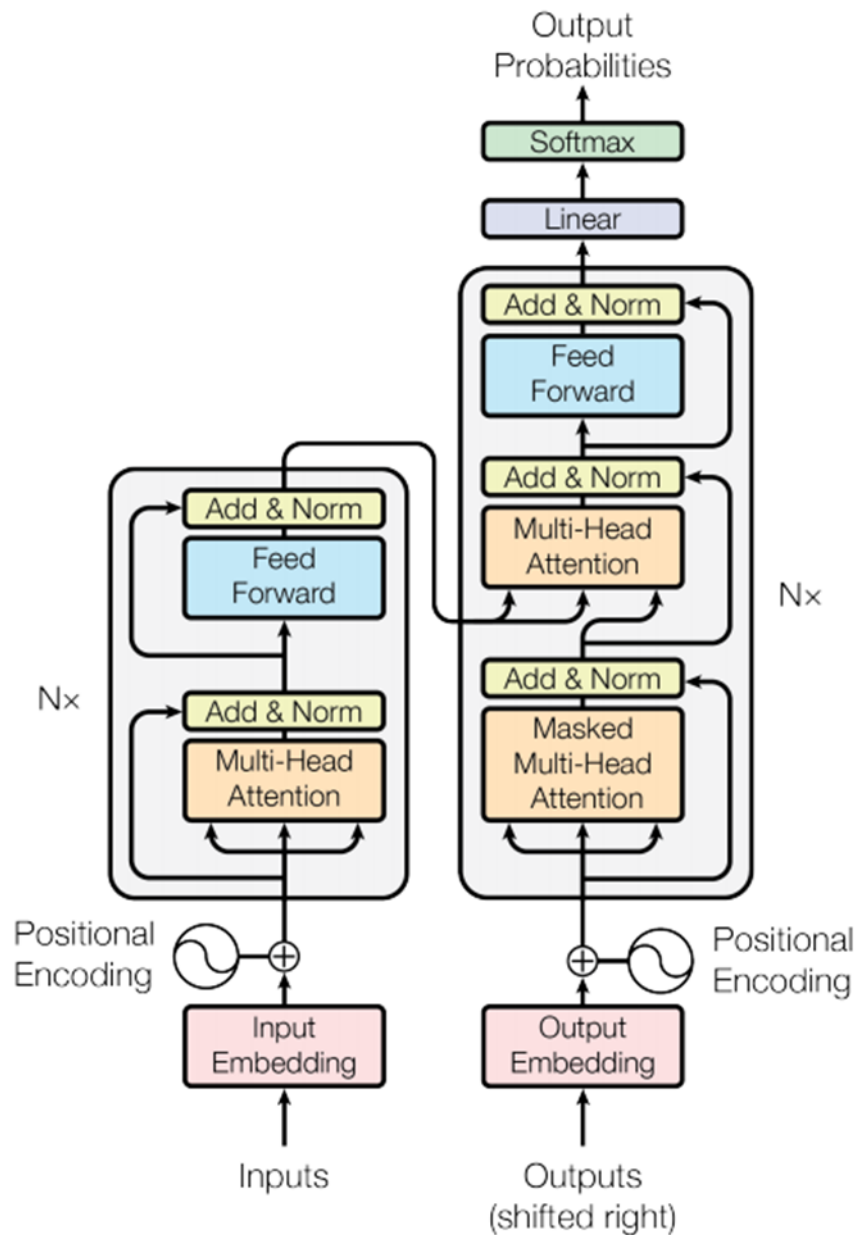
Sendo $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d \times d}$ parâmetros que serão aprendidos durante o treino da rede neural. (WU et al., 2021)

Figura 13 – Representação de uma Camada de Multi-Head Attention



Fonte: Vaswani et al. (2017)

Figura 14 – Arquitetura do Transformer



Fonte: Vaswani et al. (2017)

A figura 14 ilustra a arquitetura de um bloco *Transformer* para uma rede neural, com o funcionamento de suas camadas de *multi-head attention* demonstrado na figura 13.

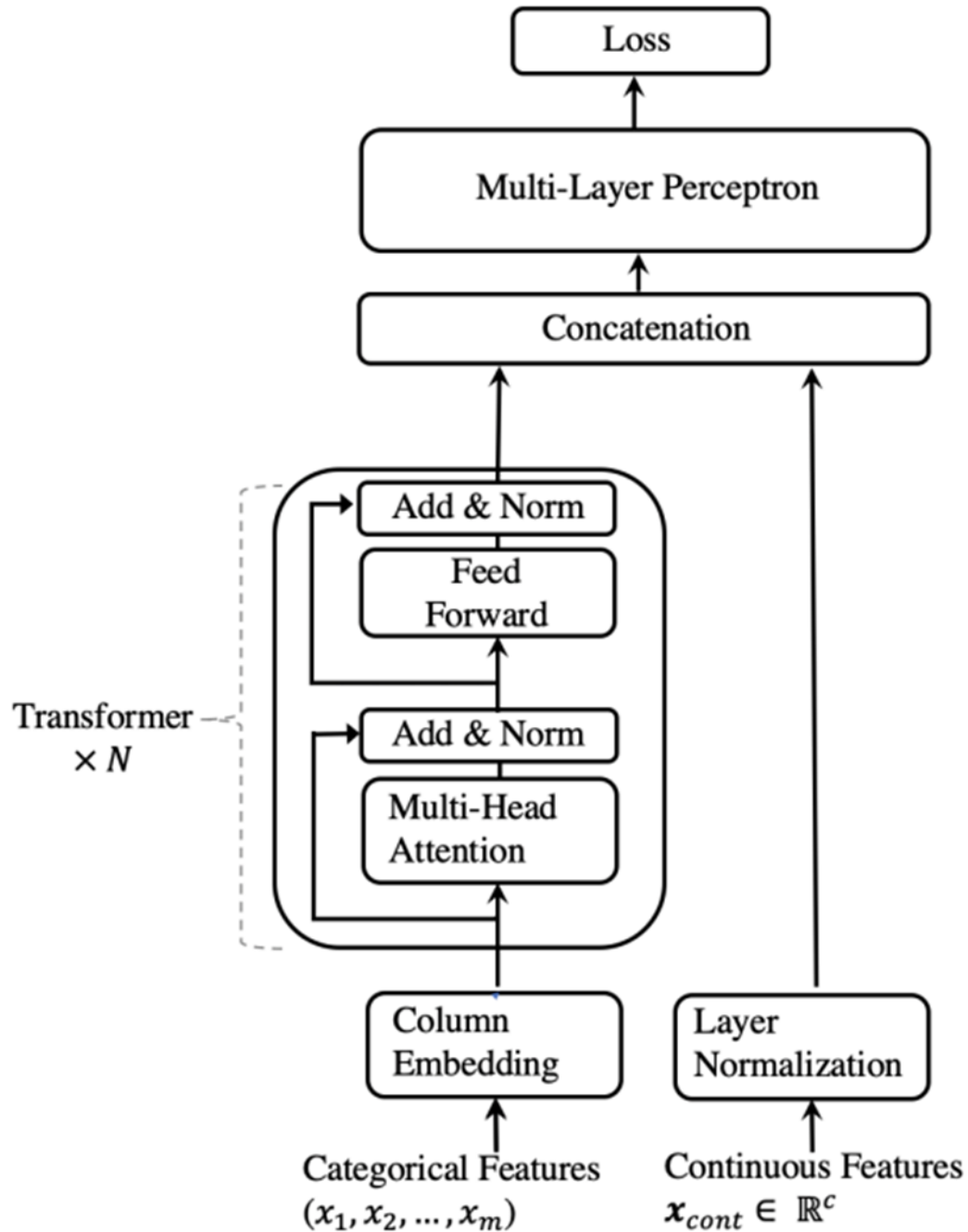
1.2.9.1 TABTRANSFORMER

Huang et al. (2020) descrevem a arquitetura do TabTransformer como uma camada que cria os *embeddings* a partir somente das variáveis categóricas, um conjunto de N camadas de *Transformers*, e no final um Perceptron de Múltiplas Camadas. Cada camada dos *Transformers*, por sua vez, consiste em uma camada de *multi-head self-attention* seguida por outra de *Feed Forward*. Sendo essa arquitetura representada na figura 15.

Nesse modelo as variáveis numéricas são normalizadas e concatenadas com o resultado

que o *Transformer* obtém a partir das variáveis categóricas; essa nova representação dos dados por sua vez é alimentada em um *Multi-layer Perceptron* que tem como seu resultado a predição do modelo final.

Figura 15 – Arquitetura do TabTransformer



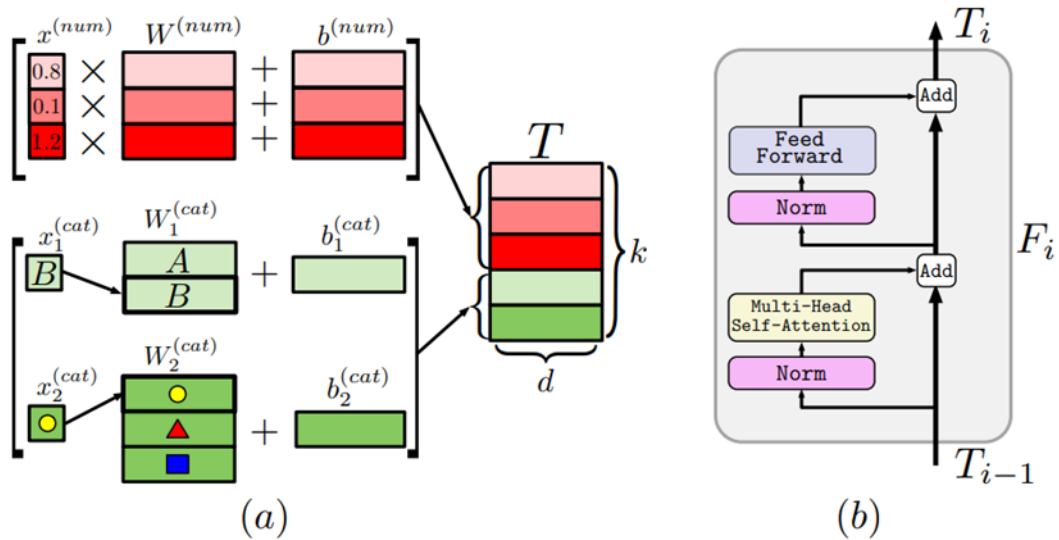
Fonte: Huang et al. (2020)

1.2.9.2 FT-TRANSFORMER

Gorishniy et al. (2021) caracterizam o FT-Transformer como uma adaptação simples do arquitetura original do *Transformer*, realizando inicialmente uma transformação de todas as variáveis (além das categóricas, como no caso anterior, ele também utiliza das variáveis

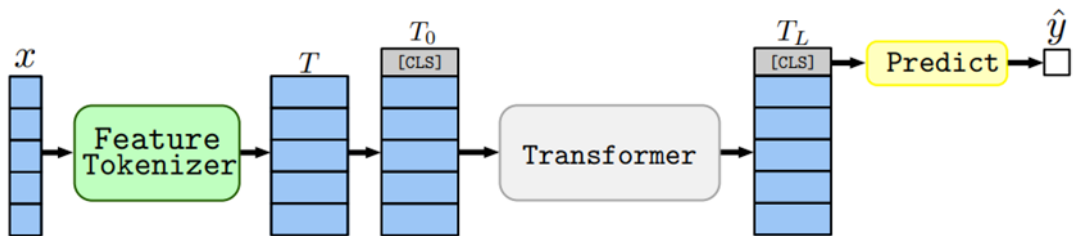
numéricas) em *embeddings* que serão processados por camadas de *Transformers*, em que cada uma delas lidará com uma variável de cada objeto, sendo que a representação resultante da última camada será a predição do modelo final, como caracterizado nas figuras 16 e 17.

Figura 16 – Feature Tokenizer



Fonte: Gorishniy et al. (2021)

Figura 17 – Arquitetura do FT-Transformer



Fonte: Gorishniy et al. (2021)

1.2.9.3 FASTFORMER

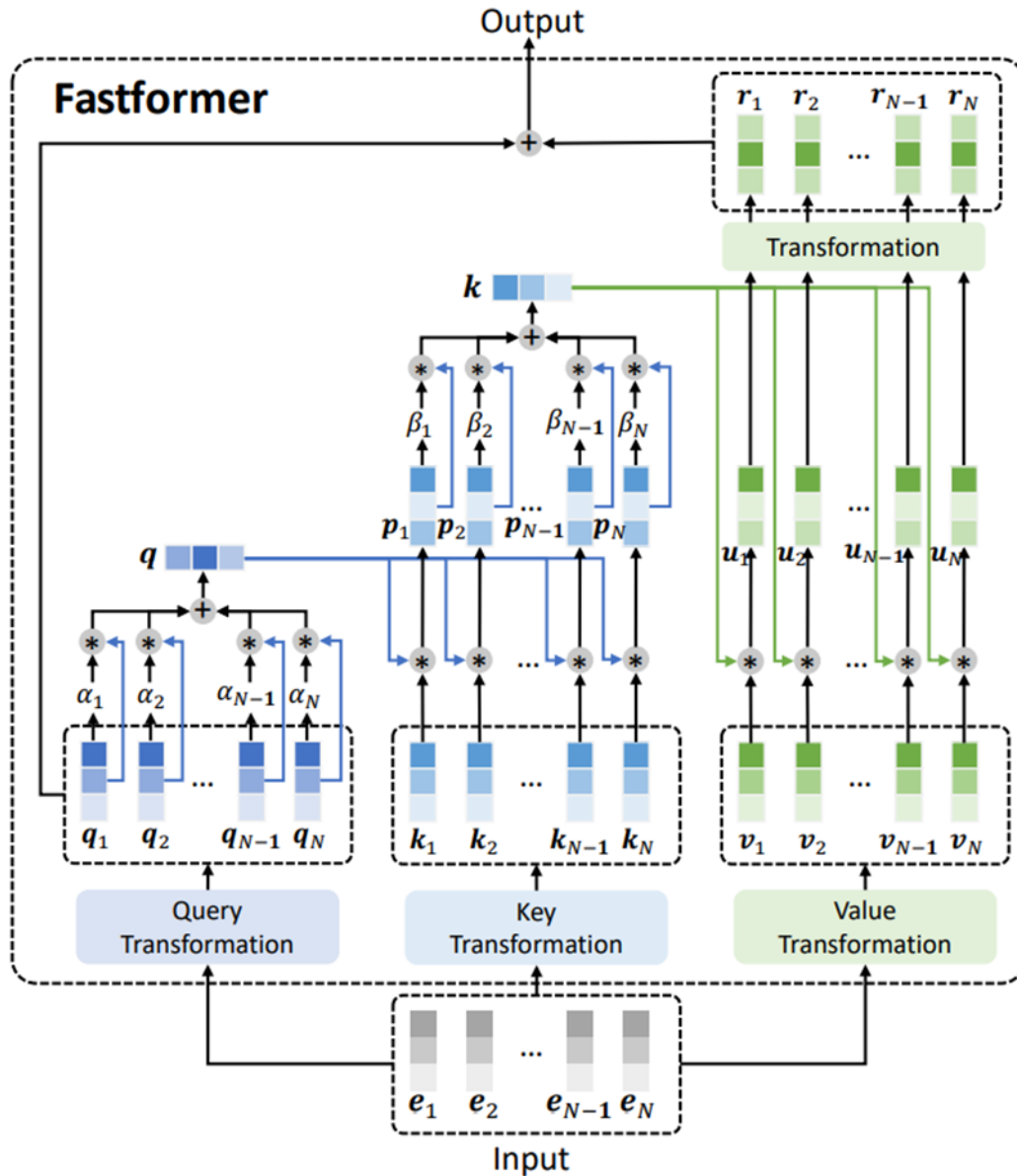
Como o processo de *self-attention* computa o produto escalar entre as representações de entrada em cada par de posições, sua complexidade computacional é quadrática em relação ao tamanho da sequência de entrada, o que torna difícil para o Transformer convencional lidar eficientemente com entradas muito extensas. (WU et al., 2021)

O FastFormer foi originalmente proposto para o Processamento de Linguagem Natural, sendo que a formulação de sua arquitetura foi motivada como, segundo Wu et al. (2021), "uma variação eficiente do *Transformer* baseada em *additive attention* que consegue atingir uma modelagem efetiva em complexidade linear".

Primeiramente usamos o mecanismo de *additive attention* para sumarizar a *query* em um vetor de *query* global. Após isso, modelamos a interação entre esse vetor e os valores-chave da Atenção através do produto de cada elemento

e resumimos esse vetor global de valores-chave através de *additive attention*. Então modelamos a interação entre os valores através do produto de cada elemento e usa uma transformação linear para aprender os valores da Atenção no contexto global. Finalmente, adicionamos a *attention query* no resultado final. (WU et al., 2021)

Figura 18 – Arquitetura do FastFormer



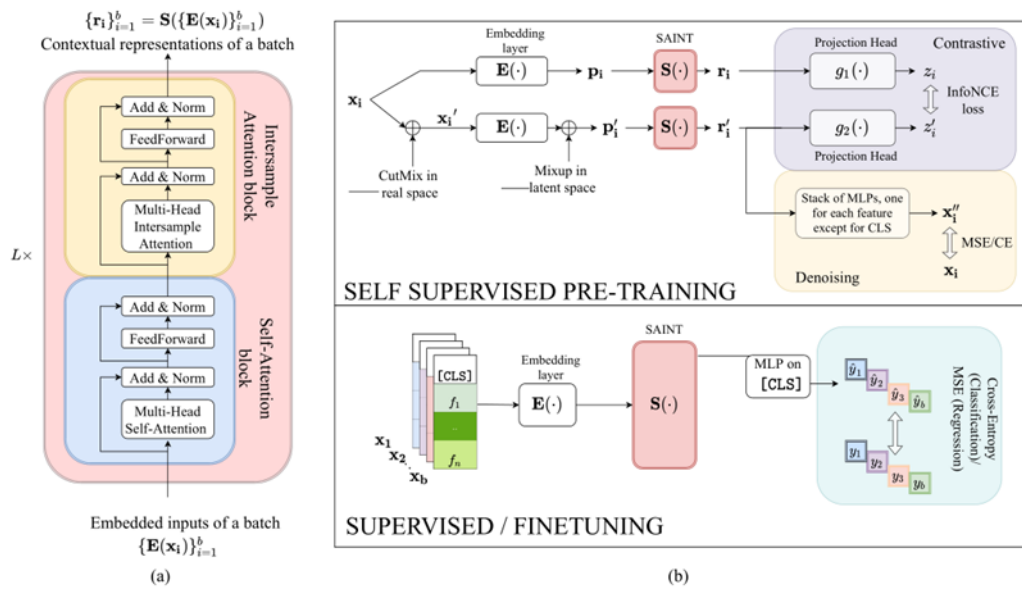
Fonte: Wu et al. (2021)

1.2.9.4 SAINT

Somepalli et al. (2021) descrevem a arquitetura do SAINT como baseada na do *Transformer* de Vaswani et al. (2017), com o modelo recebendo uma sequência de *embeddings* das variáveis explicativas e retornando uma saída contextualizada de mesma dimensão.

O SAINT é composto por L camadas idênticas, cada uma delas sendo composta por um bloco de *self-attention transformer* e um de *intersample attention transformer*. (SOMEPALLI et al., 2021)

Figura 19 – Arquitetura do SAINT

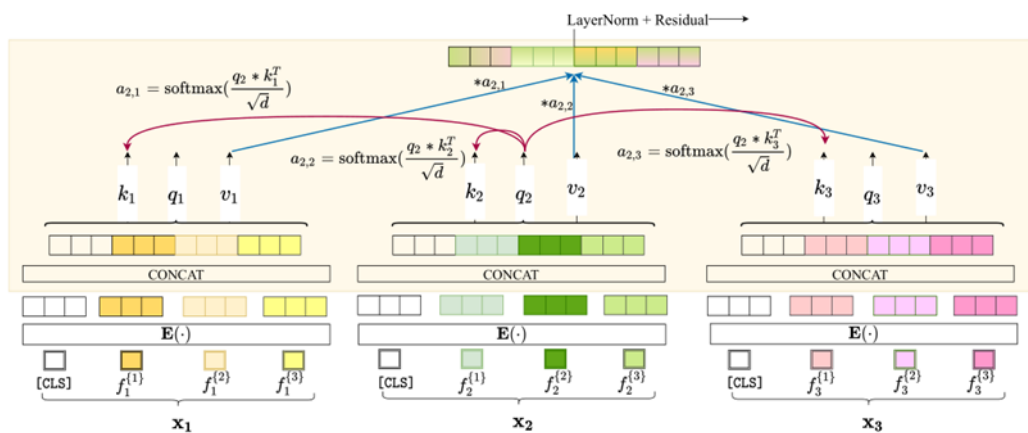


Fonte: Somepalli et al. (2021)

Somepalli et al. (2021) descrevem a Atenção Entre Amostras (*Intersample Attention*) como sendo um processo de Atenção computada para um *batch* ao invés de calcular para as variáveis de um único ponto. Assim, são concatenados os embeddings de cada variável de um dado ponto, e se computa a atenção sobre as outras amostras, de acordo com o diagrama da figura 20.

Isso permite melhorar a representação de cada ponto, ao se examinar outros pontos. Quando uma dada variável é faltante ou tem alto ruído para uma amostra, a Atenção Entre Amostras permite que o SAINT copie valores dessa variável de outros pontos, considerados similares, dentro do mesmo *batch*. (SOMEPALLI et al., 2021)

Figura 20 – Representação do Mecanismo de Atenção Entre Amostras



Fonte: Somepalli et al. (2021)

1.3 INTERPRETABILIDADE

A capacidade de corretamente interpretar a predição dada por um modelo é extremamente importante. Ela cria confiança em quem faz seu uso, provê conhecimento de como o modelo pode ser melhorado e embasa o entendimento do processo modelado. Em algumas aplicações, modelos mais simples (por exemplo, modelos lineares) são preferidos pela sua facilidade de interpretação, mesmo que possam ser menos acurados do que modelos mais complexos. (LUNDBERG; LEE, 2017)

Também demonstram sua utilidade para construir *ensembles*, modelos que mesmo não tendo a melhor performance, mas que valorizam variáveis diferentes tendem a estar menos correlacionados aos demais, melhorando a predição final do *ensemble*.

Outro uso é na detecção de vazamento de dados, quando uma variável explicativa é tida como extremamente relevante, superando por larga vantagem as outras que compõe o modelo; ela deve então ser investigada para entender qual sua relação com a variável resposta, e se contém informações que não estariam disponíveis caso o modelo fosse colocado em prática com novos dados.

Um dos métodos para auxiliar na interpretabilidade dos modelos desse estudo será o SHAP, introduzido por Lundberg e Lee (2017), que faz uso do conceito da Teoria dos Jogos, de Valores Shapley.

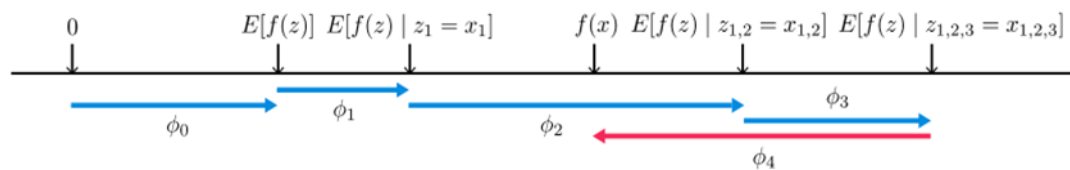
Os valores das variáveis em uma dada amostra se comportam como jogadores em um time. Os Valores Shapley nos informam sobre como distribuir a recompensa do resultado (no caso, predição) entre as variáveis. Um jogador pode ser um único valor de uma variável, por exemplo, em dados tabulares. (MOLNAR, 2022)

Uma inovação trazida pelo SHAP é que a explicabilidade dos Valores Shapley é representada de uma forma aditiva, como um modelo linear. (MOLNAR, 2022)

Para uma única amostra, a importância atribuída para uma dada variável é calculada através da média das diferenças entre os resultados que o modelo teria utilizando os diferentes subconjuntos de variáveis que ela está presente e um valor base, que consiste na média da variável resposta, uma predição que não leva em conta qualquer variável explicativa. Esse processo é então repetido para outras amostras com o intuito de se obter uma visão global sobre o impacto das variáveis, tanto individualmente quanto o de suas interações, no modelo.

Irá ser calculado tanto a média do valores SHAP para uma fração das amostras do conjunto de treino, nos dando informações sobre se cada variável tem um impacto positivo ou negativo no valor predito, quanto a média de seus absolutos, nos dizendo a importância que cada variável tem para o modelo.

Figura 21 – Representação da Influência Individual de Variáveis Descrita pelo SHAP



Fonte: Lundberg e Lee (2017)

Uma segunda técnica de interpretabilidade usada será o *Feature Permutation*, nela amostras do conjunto de teste tem os valores de suas variáveis explicativas substituídos aleatoriamente, com o de outras amostras, da mesma variável. Então é calculada a média das diferenças entre as predições que o modelo realizou para o banco de dados original e para o perturbado.

Variáveis que foram perturbadas nos subconjuntos de maior diferença em módulo são consideradas as mais relevantes.

Para todos os bancos de dados, os valores do SHAP, *Feature Permutation* e Importância Interna foram normalizados, para cada modelo, dividindo pela soma dos absolutos dos valores atribuídos a todas as variáveis.

É importante ressaltar que quando se calcula a média das importâncias para um grande número de amostras se obtém uma visão global, e não local, da interpretabilidade do modelo, então variáveis que sejam muito importantes para a predição de poucas amostras podem parecer pouco importantes para o conjunto.

1.3.1 INTERPRETABILIDADE PARA REDES NEURAIAS

Arik e Pfister (2020) descrevem a capacidade de interpretação da influência de cada variável no Tabnet através da seleção esparsa de variáveis que ele realiza. Há o emprego de múltiplos blocos de decisão que focam em processar somente um subconjunto das variáveis disponíveis a cada passo, podendo assim mensurar sua importância durante o processo de aprendizado.

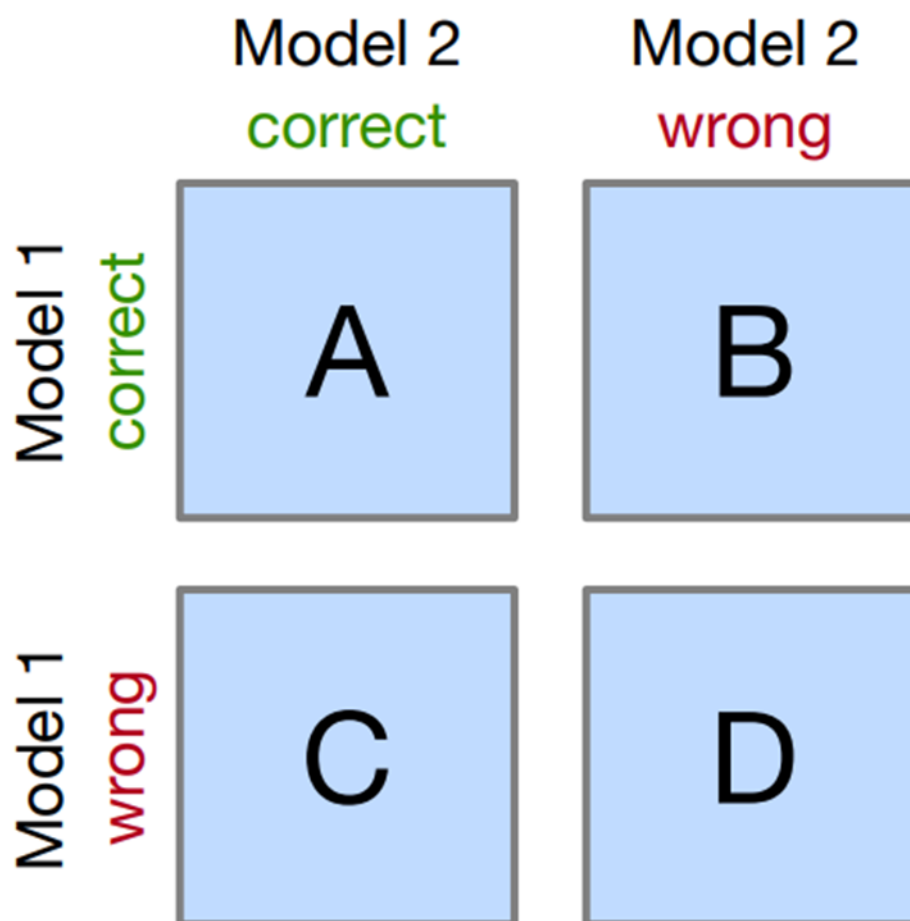
Durante o processo de *backpropagation* da rede as máscaras vão sendo atualizadas podendo omitir mais frequentemente certas variáveis, caso elas sejam consideradas menos relevantes para a predição.

Para o SAINT, Somepalli et al. (2021) interpretam a predição através do processo de Atenção produzido pelos *transformers*. Descrevedo que, em particular, quando só se usa um estágio de *Transformer* o mapa da Atenção revela todas as variáveis e quais amostras foram usadas pelo modelo para realizar a decisão.

1.4 TESTE DE HIPÓTESE PARA CLASSIFICADORES BINÁRIOS

O teste de McNemar é um teste não-paramétrico para comparação do desempenho de dois classificadores binários, Raschka (2020) descreve seu procedimento com a criação de uma matriz 2x2 em que seu primeiro elemento é a quantidade de amostras classificadas corretamente pelos dois algoritmos comparados e seu último o número de classificados erroneamente por ambos. Os elementos fora da diagonal principal são a quantidade de amostras classificadas erroneamente por somente um algoritmo.

Figura 22 – Matriz do Teste de McNemar



Fonte: Raschka (2020)

No teste de McNemar formulamos a hipótese nula como as probabilidades $P(B)$ e $P(C)$ – em que B e C se referem aos elementos fora da diagonal principal da matriz – serem iguais, em outras palavras que nenhum modelo teve um desempenho melhor que o outro. Sendo que consideramos a hipótese alternativa como a performance dos modelos não ser igual. Então a estatística do teste de McNemar pode ser computada como: (RASCHKA, 2020)

$$\chi^2 = \frac{(B - C)^2}{B + C} \quad (9)$$

A estatística do teste de McNemar, χ^2 , segue uma distribuição Qui-Quadrado com um grau de liberdade (assumindo que a hipótese nula é verdadeira e temos um número relativamente grande de elementos em B e C, por exemplo, maior do que 25). (RASCHKA, 2020)

Raschka (2020) descreve a correção de continuidade proposta por Edwards (1948) como a forma mais utilizada correntemente da estatística para esse teste. Justificando seu uso através do aumento da acurácia do teste quando se lida com frequências discretas para a distribuição Qui-Quadrado.

$$\chi^2 = \frac{(|B - C| - 1)^2}{B + C} \quad (10)$$

Na comparação feita por Dietterich (1998) o teste de McNemar apresentou uma boa performance para avaliação de classificadores, com a vantagem de não ser necessário múltiplos processos de treinamento com diferentes partições do banco de dados, processo dispendioso para métodos mais caros computacionalmente como redes neurais.

Para a avaliação de mais de dois modelos Raschka (2020) sugere a realização inicialmente do teste Q de Cochran, uma generalização do teste de McNemar, o descrevendo como similar a uma ANOVA, mas para dados pareados, não nos dizendo quais grupos testados são diferentes, mas se há alguma diferença estatisticamente significativa entre eles. Nele a hipótese nula descreve o evento que a acurácia de todos os modelos são iguais e ela segue uma distribuição Qui-Quadrado com graus de liberdade iguais ao número de modelos testados menos um.

Seja $\{C_1, \dots, C_M\}$ um conjunto de classificadores testados em algum banco de dados. Se os M classificadores não diferem em sua performance, então a seguinte estatística Q é distribuída aproximadamente como Qui-Quadrado com M-1 graus de liberdade: (RASCHKA, 2020)

$$Q = (M - 1) \frac{M \sum_{i=1}^M G_i^2 - T^2}{MT - \sum_{j=1}^n M_j^2} \quad (11)$$

Sendo G_i o número de amostras corretamente classificadas por $C_i = 1, \dots, M$; M_j o número de modelos que corretamente classificaram o j-ésimo exemplo no banco de dados de teste e T é o número total de votos corretos entre os M classificadores. (RASCHKA, 2020)

Caso a hipótese nula seja rejeitada, então serão realizados testes de McNemar *post-hoc* de cada par das redes neurais, para um nível de significância de 5% com correção de Holm-Bonferroni para múltiplos testes.

2 BANCOS DE DADOS

Com o intuito de demonstrar o desempenho de cada algoritmo em diferentes cenários foram selecionados bancos de dados de diferentes áreas, que representam problemas complexos, com dimensões muito distintas e que possuem tarefas diferentes para a predição.

O primeiro banco de dados (HOPPEN, 2017) consiste em informações anonimizadas de consultas médicas do Sistema Único de Saúde no estado do Espírito Santo, nos anos de 2015 e 2016. As variáveis explicativas são as datas do agendamento e da consulta, onde podemos extrair qual dia da semana e do mês que ocorrem, além da diferença entre as datas, também há características do paciente, como seu sexo, idade, recebimento ou não do programa de assistência Bolsa-Família e possíveis problemas de saúde, tendo uma variável resposta binária que trata se o paciente atrasou ou não na consulta agendada.

São 110527 amostras (destas 20,2% correspondem a atrasos), cada uma representando uma consulta, que após o pré-processamento dos dados possuíam cada 40 variáveis explicativas.

O segundo banco de dados (AGRAWAL, 2017) trata de diamantes avaliados pela joalheria Tiffany & Co's. Após o pré-processamento são 53920 amostras que possuem 26 variáveis explicativas que tratam de características da pedra preciosa, como seu quilate, dimensões, corte, cor e categoria de qualidade que foi classificada, tendo como variável resposta o preço que ela foi avaliada em dólares americanos, ajustados pela inflação do período.

O terceiro banco de dados (MCCAULEY, 2018) descreve 31892 colisões de pares de múons e antimúons, partículas subatômicas instáveis e similares ao elétron, no *Large Hadron Collider*, disponibilizado abertamente pelo CERN. Após o pré-processamento foram obtidas 17 variáveis explicativas que representam características de cada partícula, como sua carga e energia, além de características da colisão como o ângulo entre elas. A variável resposta é a soma da massa invariante (também chamada de massa de repouso), na unidade GeV, medida que independe dos movimentos no sistema e do referencial do observador, de ambas as partículas.

O quarto banco de dados é a associação de dois publicados por Hamidieh (2018) que tratam de 21263 supercondutores com composições químicas distintas que tiveram sua temperatura crítica, definida como a maior temperatura em Kelvin onde ele apresenta o fenômeno da supercondutividade, medidas. Há um grande número de variáveis explicativas, contendo informações sobre o número de elementos presentes, quais deles compõe o material, além de funções, como média ponderada e desvio padrão, das propriedades físicas dos átomos presentes, como o calor de fusão, raio atômico e condutividade térmica. Intenciona-se que esse banco de dados teste a capacidade de cada modelo para a seleção de variáveis, dado o seu alto volume, além da capacidade de evitar o sobreajuste.

O quinto banco de dados disponibilizado por Lyon et al. (2016), trata de 17898 observações coletadas pelo *High Time Resolution Universe Survey (South)* tendo sua variável resposta separada por físicos em duas classes: ruído, podendo ser causado pela radiação cósmica ou interferência eletromagnética, e emissões eletromagnéticas das estrelas de neutrôn da categoria Pulsar.

Ele apresenta um profundo desbalanceamento entre as classes, tendo o ruído como mais de 90% das situações, e tem como variáveis explicativas estatísticas, como a média, desvio-padrão, assimetria e curtose em excesso, dos dois tipos de medições realizadas, o Perfil de Pulso

Integrado, curva fruto da integração dos sinais periódicos captados, e da Curva DM-SNR, que descreve o atraso medido entre as diferentes frequências eletromagnéticas captadas. Para esse banco de dados não foram realizadas técnicas de reamostragem dos dados, é desejado observar como cada algoritmo se comporta e se consegue evitar a solução trivial, em que é predito para todas as amostras a classe majoritária.

O sexto banco de dados (KOKLU; OZKAN, 2020) envolve 13611 diferentes feijões, com sua variável resposta tratando de qual das 7 espécies ele pertence, que foram fotografados e tiveram suas *features* extraídas por um processo de visão computacional, resultando nas suas 16 variáveis explicativas referentes a sua morfologia, como sua área, perímetro, excentricidade e comprimento em diferentes seções.

O sétimo banco de dados, disponibilizado por Moro, Cortez e Rita (2014), abrange campanhas de *telemarketing* de um banco português no período entre Maio de 2008 e Novembro de 2010. Com o intuito de acelerar o processamento dos modelos foi realizado um processo de amostragem da classe majoritária, aquela em que não houve sucesso na venda do produto e que corresponde a cerca de 90% das amostras, para reduzir o *dataset* para pouco acima de 25% de seu tamanho original.

Resultando em 10300 amostras, cada uma com 45 variáveis explicativas que envolvem dados do cliente como sua idade e profissão, além de informações relativas às campanhas de publicidade, como se houve contato com esse mesmo cliente em campanhas anteriores, se ela obteve êxito, além do dia da semana e mês que o contato atual ocorreu.

O oitavo banco de dados é produto de uma junção de dois publicados por Cortez et al. (1998), ambos tratam de vinhos portugueses, originalmente separados entre brancos e tintos, da Denominação de Origem Controlada Vinho Verde. Tendo como variáveis explicativas características físico-químicas, como seu pH, densidade, teor alcóolico, açúcar residual e concentração de cloretos, e sua variável resposta é numérica e descreve a mediana das notas que eles receberam de críticos em testes cegos. No total são 6497 amostras e 12 variáveis explicativas.

O nono banco de dados (NASH et al., 1994) diz respeito originalmente a 4177 moluscos da espécie *Haliotis*, comumente conhecida como Abalone. O padrão-ouro para a determinação da idade, em anos, desse animal é somar 1,5 ao número de anéis vistos, sob um microscópio, internos a sua concha, trabalho de análise demorado e que resulta na morte do animal, o que motiva a busca por formas de determinar esse valor evitando esse processo. Nesse banco de dados temos como variável resposta o número de anéis que foram contados em cada animal e para as variáveis explicativas temos as dimensões físicas, sexo e pesos observados.

O décimo banco de dados (CAMPOS et al., 2000) trata de 2126 exames de cardiocografia, procedimento comum para avaliar a saúde do feto no último trimestre de gestação, avaliados por três médicos que classificaram seus resultados como saudável, suspeito e patológico. Suas 20 variáveis explicativas são funções referentes à frequência cardíaca do feto e contrações uterinas durante o exame.

O décimo primeiro banco de dados foi disponibilizado por Yeh (1998) e trata de 1030 amostras de concreto em que foram medidas a resistência à compressão, sua variável resposta, e tem como suas 8 variáveis explicativas a concentração de substâncias adicionadas no seu preparo, além da quantidade de dias que se passaram entre o teste e sua fabricação.

O décimo segundo banco de dados (GRISONI et al., 2015) (GRISONI et al., 2016) contém 9 fatores de descrição química de 779 moléculas, e sua variável resposta é o Fator de Bioconcentração, que mensura o acúmulo de uma dada substância química em um ser vivo

quando absorvida pela água, medido na escala negativa logarítmica.

O décimo terceiro banco de dados (ATLAS Collaboration, 2014), é o único dos em estudo que envolve dados simulados, sendo originalmente proposto, em colaboração com o CERN, como uma competição na plataforma Kaggle. Ele envolve um processo de classificação binária, onde as amostras da classe positiva da variável resposta são dados de colisões onde seriam gerados Bósons de Higgs, segundo as propriedades que os pesquisadores acreditavam que essa partícula teria. Após o pré-processamento ele contém 237985 amostras, cada uma com 30 variáveis explicativas, que descrevem os jatos de partículas subatômicas gerados, além da massa estimada da partícula que possivelmente é o Bóson de Higgs.

Tabela 1 – Informações dos Bancos de Dados Após o Pré-Processamento

Banco de Dados	Tamanho Amostral	Variáveis Explicativas	Tarefa
Bóson de Higgs	237985	30	Classificação (Binária)
Atrasos SUS	110527	40	Classificação (Binária)
Diamantes	53920	26	Regressão
Muóns	31892	17	Regressão
Supercondutores	21263	158	Regressão
Pulsares	17898	8	Classificação (Binária)
Feijões	13611	16	Classificação (7 Classes)
Marketing Banco	10300	45	Classificação (Binária)
Vinhos	6497	12	Regressão
Abalone	4171	9	Regressão (Contagem)
Cardiotocografia	2126	20	Classificação (3 Classes)
Concreto	1030	8	Regressão
Bioconcentração	779	9	Regressão

2.1 SEPARAÇÃO DOS DADOS

Os dados foram separados aleatoriamente em conjuntos de treino, validação e teste. O primeiro será usado para treinar cada um dos modelos com diferentes seleções de hiperparâmetros, com seus resultados comparados no conjunto de validação. Após a melhor combinação de hiperparâmetros ser selecionada no conjunto de validação, cada modelo escolhido terá seu desempenho medido no conjunto de testes.

Com o intuito de que os algoritmos possam ser treinados e testados em partições dos dados com características similares aos de sua totalidade, para os *datasets* de classificação foi realizada uma estratificação que mantém uma proporção similar das classes da variável resposta em cada conjunto. De forma similar, para os de regressão as amostras foram separadas de acordo com a proporção dos quantis dos valores que a variável resposta possa assumir.

Todo o processamento dos dados, como sua normalização, foi feito de forma separada para cada conjunto, evitando assim que informações que estejam em no conjunto de validação ou teste possam ser utilizadas pelos algoritmos durante o treino, ocorrência denominada de vazamento de dados (*data leakage*).

Em todos os bancos de dados foram separados 15% das amostras para os conjuntos de teste e validação, sendo que as 70% restantes foram para o de treino.

Para o conjunto de dados dos vinhos, Pulsares, supercondutores e feijões o processo de validação ocorreu com 3 conjuntos de validação e treino diferentes, para o do concreto e bio-concentração foram 10 conjuntos e para o do Abalone e cardiocardiografia 5, processo conhecido como *K-Fold Cross-Validation*, para todos os outros foi feito somente 1 *fold* de validação. A métrica escolhida foi medida para cada separação e então calculada a média aritmética.

Após a escolha da melhor combinação de hiperparâmetros, os conjuntos de treino e validação serão unidos e o modelo escolhido será treinado com a totalidade deles (85% dos dados), com exceção das redes neurais, em que uma parcela do número de amostras na razão de $1/\sqrt{2}$ será separada para o *early-stopping*, como sugerido por Amari et al. (1995).

2.2 TRATAMENTO DOS DADOS

Alguns critérios para a exclusão de amostras e variáveis foram decididos antes da exploração dos dados, sendo regras gerais para o rigor do estudo:

- Serão excluídas as amostras que apresentam a variável resposta faltante, pois o estudo não se propõe a realizar Aprendizado Auto-Supervisionado, e aquelas que apresentam as observações para todas as variáveis explicativas faltantes, pois não há qualquer informação que possa ser capturada para a predição.
- Serão excluídas as variáveis explicativas com variância igual zero, dado que é impossível entender qualquer relação que ela possa possuir com a variável resposta, e também as que apresentam Correlação de Pearson absoluta igual a um com outra explicativa, dado que mantê-las no modelo levaria a um maior custo computacional e problemas de multicolinearidade sem que exista qualquer nova informação que elas possam agregar.
- Por fim, serão excluídas as variáveis explicativas originais com *Variance Inflation Factor* infinito, pois elas são somente uma combinação linear de outras já presentes.

Todas as variáveis categóricas foram transformadas em múltiplas variáveis binárias que tratam do pertencimento ou não da amostra em dada categoria, processo conhecido como *One-Hot Encoding* ou *Dummy Encoding*.

Não foi feita seleção de variáveis além das descritas, foi desejado observar como cada algoritmo pode realizar internamente esse processo, além do impacto que um alto número de variáveis pode causar na qualidade da predição, como a propensão de cada modelo ao sobreajuste.

Para as redes neurais, com exceção do MLP, qualquer variável com menos de 100 valores únicos foi tratada como categórica para o efeito da criação de seus *embeddings*.

A Padronização dos dados foi feita da forma que, para cada conjunto separadamente, os valores de cada variável explicativa foram subtraídos de sua média (centralização) e divididos por seu desvio padrão. Resultando em novas variáveis que possuem uma mesma escala, com média zero e desvio-padrão igual a 1.

$$X_{padronizado} = \frac{X - \mu}{\sigma} \quad (12)$$

Para o banco de dados do SUS, foi desconsiderada a variável que trata do bairro do paciente, devido ao grande número de categorias e baixo número de amostras para a maior parte delas, além dos algoritmos em estudo não levarem em conta dependências espaciais.

Também houve um agrupamento das variáveis que tratam das idades dos pacientes e número de deficiências, devido ao baixo número de paciente com múltiplas deficiências e do fato que o banco de dados reporta pacientes grávidas com idades negativas, criando a necessidade de desconsiderar a idade desses paciente e criar uma nova variável binária que indica gravidez.

Para o banco de dados dos diamantes foram excluídas as 20 amostras que as dimensões espaciais da pedra eram iguais a zero e também foi feita uma transformação da variável resposta tomando o logaritmo natural dela, devido a ela ser estritamente maior do que 1 e com extrema assimetria para a direita.

O banco de dados do Abalone foi trazido para a escala original, multiplicando todas as variáveis contínuas por 200, antes de sua padronização. Também foram retiradas 6 amostras que possuíam valores incoerentes, como a altura igual a zero ou peso total menor que o peso de uma de suas partes.

Para o banco de dados dos feijões, para as redes neurais com exceção do MLP, a variável explicativa *Compactness* foi arredondada para a segunda casa decimal, garantindo que pelo menos uma das variáveis do banco tenha menos do que 100 valores únicos e que possam ser criados os embeddings necessários para o pleno funcionamento desses algoritmos. De forma similar, para o banco de dados das Pulsares, a variável explicativa que trata da curtose em excesso do Perfil Integrado foi arredondada para a primeira casa decimal.

Para o banco de dados do marketing do banco português foram desconsideradas as variáveis macroeconômicas incluídas somente na última versão da base de dados.

Para o banco de dados dos vinhos foi criada uma nova variável binária que descreve se o vinho é tinto ou branco.

Para o banco de dados do fator de bioconcentração foram consideradas como variáveis explicativas somente as variáveis teóricas obtidas antes da experimentação.

Para o banco de dados do Bóson de Higgs, devido ao seu tamanho extremo inicial, sendo mais de 800000 amostras, foi decidido estudar uma parcela do problema, a qual durante as colisões foram gerados dois ou mais jatos de partículas subatômicas, pois nesse cenário todas as variáveis presentes, como o ângulo entre os jatos e suas massas, estão definidas.

Os valores faltantes da massa estimada da partícula candidata a Bóson de Higgs foram substituídos por zero.

3 ESTUDOS

Intenciona-se que em cada experimento seja realizada uma comparação nas métricas de cada modelo com os demais e também como que cada um deles tem uma mudança na qualidade de predição em cada situação apresentada.

Em cada um dos experimentos também serão ajustados classificadores triviais para os dados, eles servirão como uma medida da performance da sorte para cada um deles. Todos eles não utilizarão as variáveis explicativas para sua predição, somente farão uso da variável resposta no conjunto de teste. Para a regressão, o classificador trivial irá prever, para todas as amostras, a média da variável resposta, e para a classificação, ele irá prever para todas as amostras a razão entre as classes ou a classe mais comum.

Para os modelos de classificação binária, nas métricas que tratam da escolha de uma categoria, como a acurácia e *Score F1*, a fronteira de decisão foi escolhida através da maximização da estatística J de Youden (YOU DEN, 1950), que representa a maximização da soma da sensibilidade e especificidade do modelo, trazendo um equilíbrio entre o falso positivo e falso negativo para cada classe. Devido a isso toda métrica de decisão foi balanceada pela prevalência de cada classe.

Para os bancos de dados de classificação, todas as classes tiveram durante o treino pesos iguais, não foi considerado um conhecimento técnico específico para cada situação que indicaria custos diferentes para cada tipo de erro e demandaria penalizações diferentes. De forma similar, nos problemas de regressão, o erro quadrático médio, utilizado para o treino, seleção de hiperparâmetros e *early-stopping* das redes, é por definição simétrico, não penalizando de forma diferente erros superiores ou inferiores ao valor real.

As redes neurais, com exceção do *Multilayer Perceptron*, foram realizadas utilizando o *framework Pytorch-Widedeep* (ZAURIN; MULINKA, 2021). O *Relaxed Lasso* através do pacote criado por Vial e Estermann (2021). Os métodos de *Gradient Boosted Trees* foram através dos pacotes implementados por seus idealizadores. Todos os outros modelos foram realizados através do pacote *Scikit-Learn* (PEDREGOSA et al., 2011).

Todas as redes neurais, com exceção do MLP, tiveram seus tempos de processamento medidos sendo realizados em uma GPU, todos os outros métodos em uma CPU.

3.1 SELEÇÃO DOS HIPERPARÂMETROS

A busca da melhor combinação de hiperparâmetros foi realizada através do pacote de otimização Bayesiana Optuna (AKIBA et al., 2019), para todos os modelos foram testadas 100 combinações viáveis de hiperparâmetros (erros que impediram as redes neurais de começarem a ser treinadas foram desconsiderados para o total de 100 combinações testadas) e escolhida aquela com a melhor métrica: A menor perda logarítmica no caso dos bancos de dados de classificação e o menor erro quadrático médio nos de regressão.

Os intervalos dos hiperparâmetros foram decididos através de considerações práticas e seguindo o sugerido nos artigos originais de cada modelo.

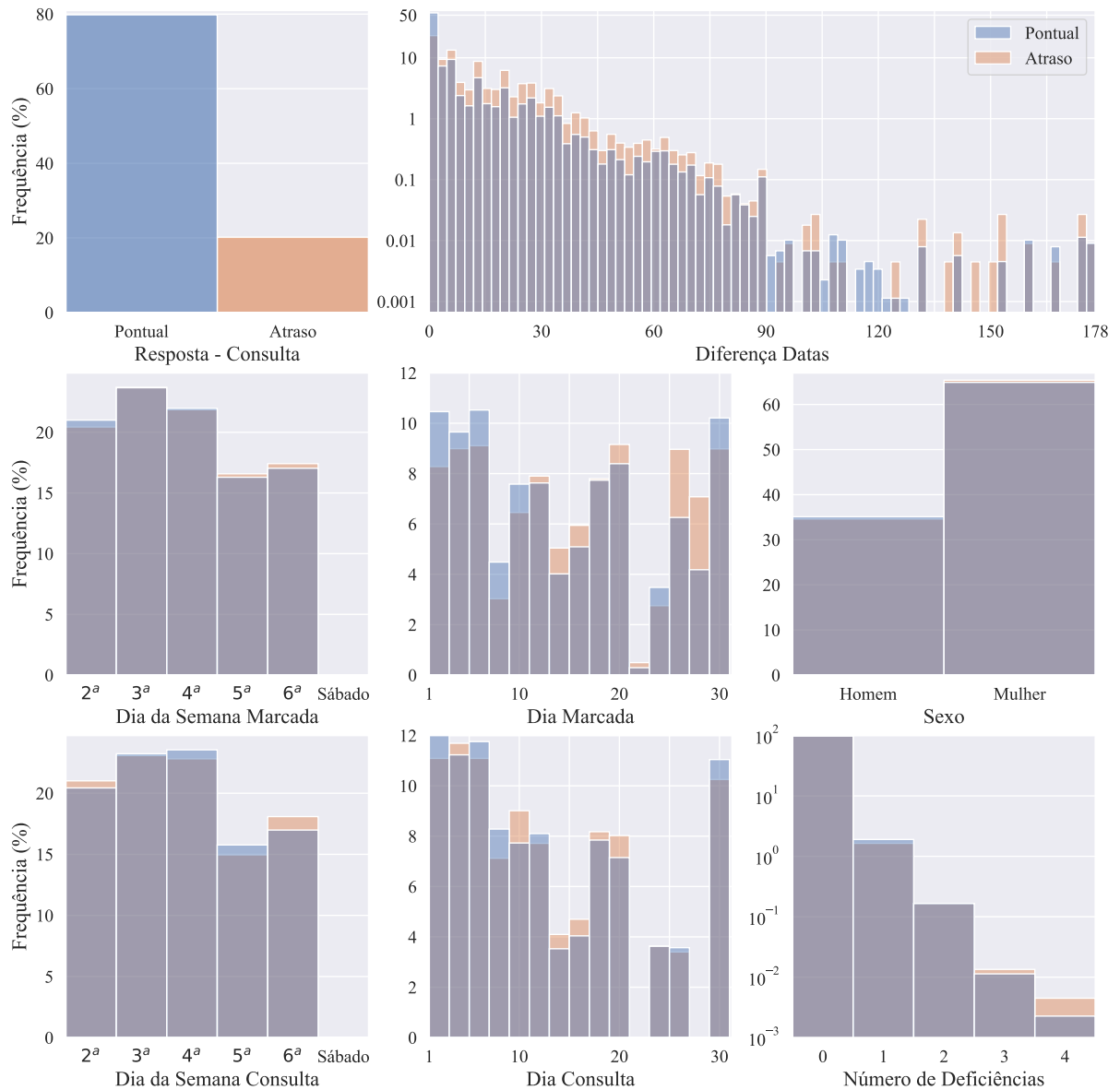
Durante a seleção de hiperparâmetros as redes neurais tem o critério de parada (*early stopping*) definido como o primeiro de dois eventos: 10 épocas consecutivas sem melhoria na função de perda igual ou superior a 10^{-4} no conjunto de validação ou 100 épocas totais de treinamento, com exceção dos bancos de dados dos atrasos no SUS, Bóson de Higgs e dos supercondutores, que devido as suas extensões, foi utilizado um limite de somente 50 épocas.

Para os modelos finais, o critério de parada é idêntico para todos os *datasets*, o primeiro evento de 100 épocas de treinamento ou 10 épocas consecutivas sem melhoria maior ou igual a 10^{-6} na função de perda no conjunto de validação.

4 EXPERIMENTOS

4.1 ATRASOS NO SISTEMA ÚNICO DE SAÚDE

Figura 23 – Histogramas das Variáveis para a Base de Dados dos Atrasos no SUS



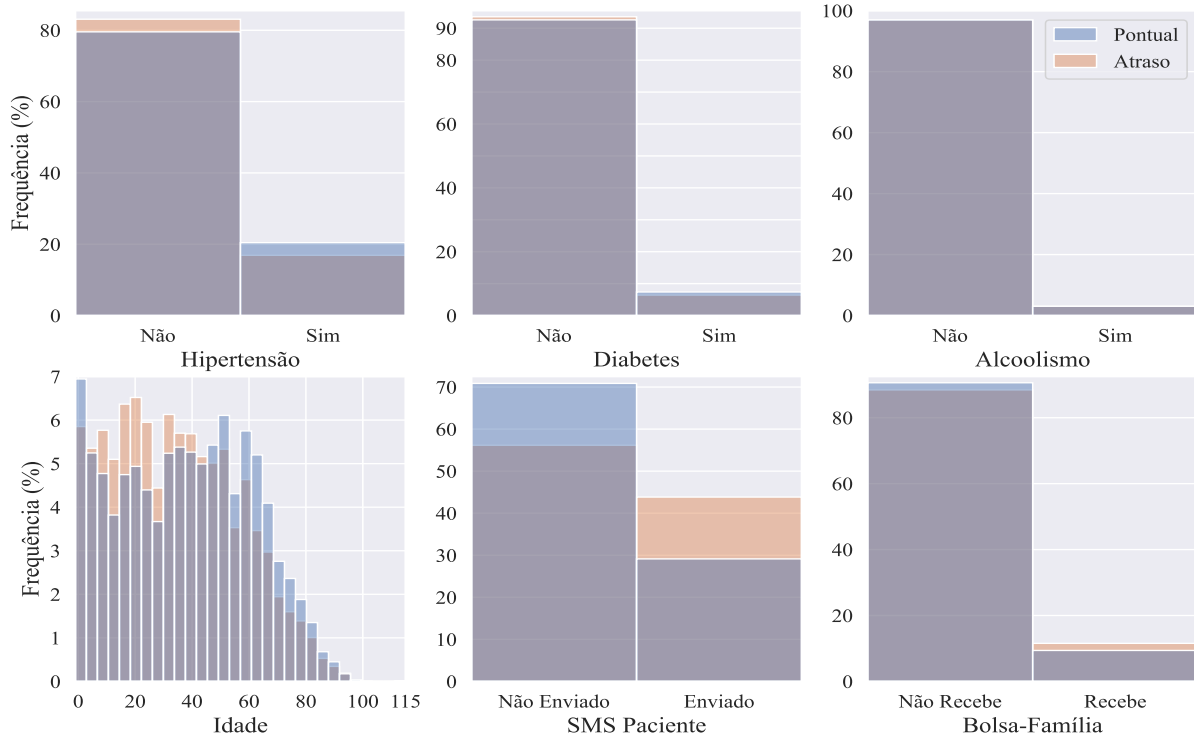
Fonte: Figura do Autor

A variável Diferença Datas denota o número de dias entre a data que a consulta foi marcada e a que ela ocorreria.

A variável SMS Paciente é binária e trata-se de se o paciente foi lembrado por SMS da data da consulta. Ela pode ser um exemplo de seleção adversa, devido ao fato de não serem todos os pacientes que tiveram a mensagem enviada, podendo ter sido privilegiados pacientes com

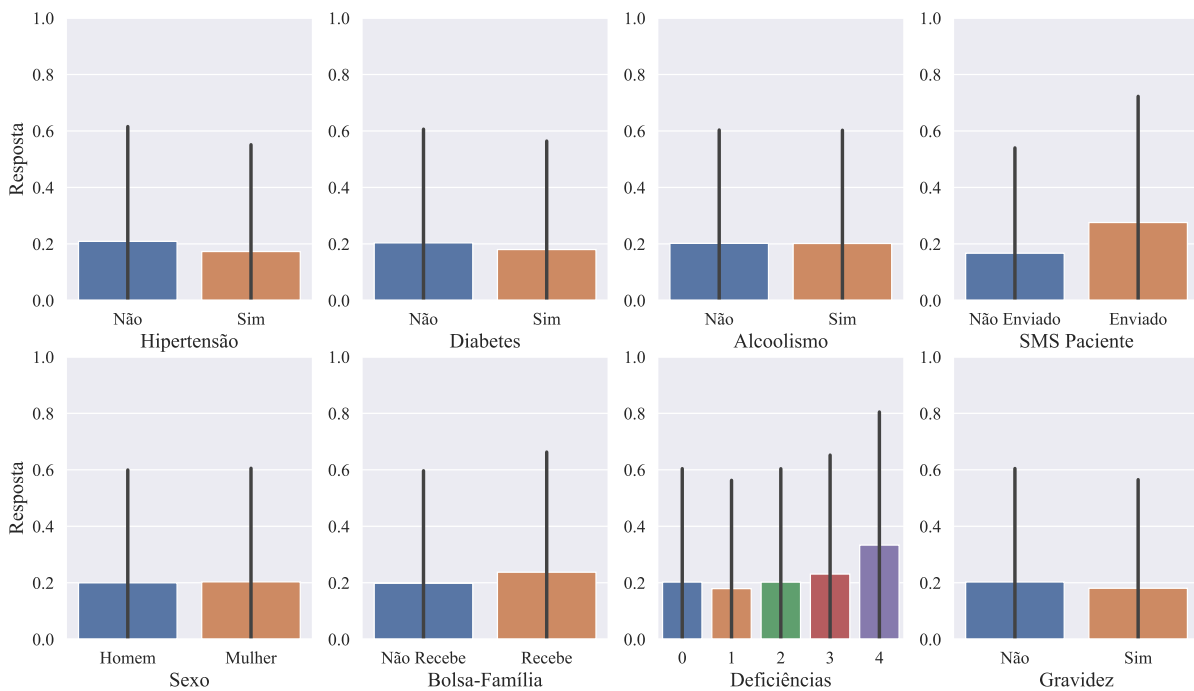
histórico de atraso, o que pode levar a que esse grupo tenha inicialmente uma probabilidade de atraso futuro maior do que a média da população.

Figura 24 – Histogramas das Variáveis para a Base de Dados dos Atrasos no SUS



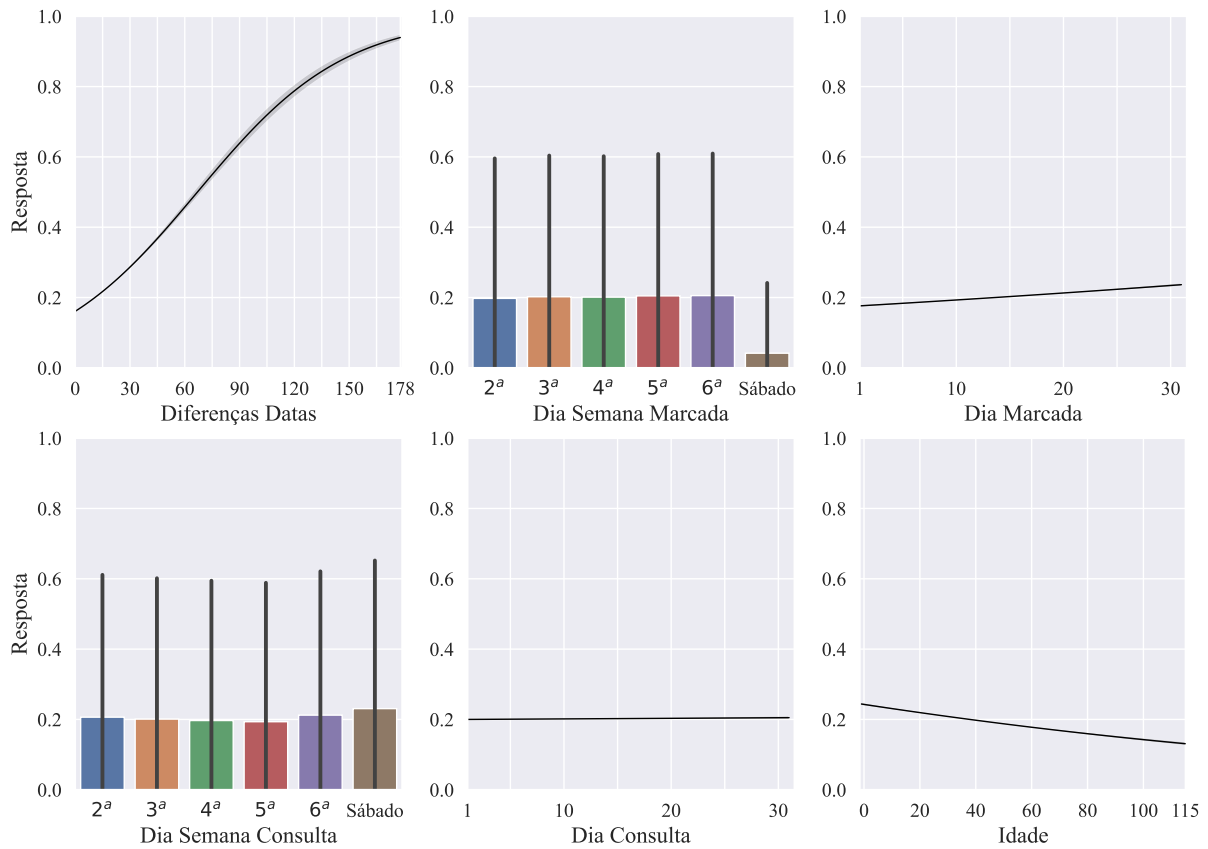
Fonte: Figura do Autor

Figura 25 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Atrasos no SUS



Fonte: Figura do Autor

Figura 26 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Atrasos no SUS



Fonte: Figura do Autor

Para todos os bancos de dados, os melhores valores em cada métrica foram ressaltados.

Tabela 2 – Métricas para a Base de Dados dos Atrasos no SUS

Modelo	Acurácia Balan- ceada	F0,5	F1	F2	AUROC	Recall	Precisão	AUPR	Log Loss	Brier Score	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,631	0,719	0,677	0,651	0,660	0,644	0,755	0,303	0,482	0,155	0,192	0,215	2,7	16
Ridge	0,630	0,720	0,680	0,655	0,659	0,648	0,754	0,303	0,482	0,154	0,192	0,214	2,1	12
Elastic Net	0,631	0,719	0,677	0,651	0,660	0,644	0,755	0,303	0,482	0,155	0,192	0,215	2,8	16
KNN	0,658	0,693	0,601	0,558	0,706	0,562	0,789	0,336	0,461	0,149	0,188	0,255	0,9	5
SVM Linear	0,632	0,715	0,667	0,638	0,662	0,631	0,757	0,304	0,482	0,155	0,188	0,215	5,2	30
Decision Tree	0,668	0,687	0,581	0,535	0,720	0,544	0,806	0,351	0,447	0,145	0,192	0,277	0,2	1
Random Forest	0,673	0,705	0,616	0,574	0,728	0,577	0,798	0,360	0,444	0,145	0,209	0,280	137,1	797
Extra Trees	0,674	0,712	0,629	0,589	0,732	0,591	0,795	0,367	0,445	0,144	0,215	0,280	166,5	968
XGBoost	0,677	0,715	0,635	0,595	0,729	0,596	0,796	0,364	0,444	0,144	0,220	0,285	34,9	203
LightGBM	0,677	0,708	0,618	0,576	0,729	0,580	0,802	0,367	0,444	0,144	0,214	0,287	7,5	44
CatBoost	0,675	0,709	0,623	0,582	0,726	0,585	0,798	0,367	0,447	0,145	0,214	0,283	116,9	680
MLP	0,666	0,710	0,632	0,593	0,721	0,593	0,788	0,355	0,450	0,146	0,208	0,267	42,3	246
ResNet	0,677	0,708	0,618	0,576	0,731	0,580	0,801	0,364	0,444	0,144	0,214	0,286	555,9	3232
TabNet	0,668	0,684	0,574	0,528	0,720	0,538	0,808	0,354	0,451	0,146	0,190	0,277	552,3	3211
SAINT	0,671	0,704	0,614	0,571	0,721	0,575	0,797	0,351	0,461	0,151	0,206	0,276	1632,4	9491
FastFormer	0,676	0,717	0,639	0,600	0,731	0,601	0,794	0,368	0,443	0,144	0,221	0,282	263,6	1533
FT-Transformer	0,675	0,712	0,629	0,589	0,732	0,591	0,796	0,367	0,444	0,144	0,216	0,282	514,6	2992
TabTransformer	0,662	0,664	0,540	0,492	0,716	0,509	0,817	0,345	0,460	0,147	0,175	0,275	369,4	2148
Trivial	0,5	0,664	0,708	0,760	0,5	0,798	0,637	0,601	0,503	0,161	0	0	-	-

É interessante notar na tabela 2 que todos os algoritmos demonstraram dificuldade em obter bons resultados, com os métodos de *gradient boosting trees* alcançando algumas das

melhores métricas com um rápido tempo de processamento, um padrão que também pode ser percebido nos próximos bancos de dados.

Foi escolhido representar as importâncias atribuídas para somente uma fração das variáveis explicativas.

Tabela 3 – SHAP Médio para a Base de Dados dos Atrasos no SUS

Modelo	Diferença Datas	SMS Paciente	Dia Marcada	Idade 18-24	Idade 13-17	> 2 Deficiências	Idade 65-69	Idade 25-29	Idade 60-64	Bolsa-Família	Consulta 2ª-Feira	Mulher
LASSO	0,388	0,180	0,072	0,071	0,022	0,000	0,025	0,040	0,004	0,023	0,010	-0,004
Ridge	0,359	0,173	0,069	0,065	0,019	0,004	0,025	0,037	0,002	0,021	0,005	-0,006
Elastic Net	0,382	0,179	0,071	0,070	0,021	0,001	0,025	0,039	0,004	0,023	0,010	-0,003
KNN	-0,064	0,145	0,165	-0,005	0,012	-0,001	-0,028	-0,015	-0,038	0,002	0,008	-0,013
SVM Linear	0,280	0,098	0,042	0,048	0,012	0,001	0,013	0,028	-0,003	0,014	-0,105	-0,004
Decision Tree	-0,632	-0,080	0,003	0,005	0,004	0,000	-0,014	0,001	-0,022	0,023	-0,023	0,004
Random Forest	-0,594	-0,089	-0,072	0,019	0,007	-0,000	-0,016	0,004	-0,019	0,018	-0,010	0,009
Extra Trees	-0,583	-0,112	-0,062	0,009	0,005	-0,000	0,007	0,002	-0,003	0,003	-0,019	0,003
XGBoost	-0,557	-0,082	-0,027	0,031	0,017	-0,000	-0,017	0,017	-0,022	0,024	-0,001	0,009
LightGBM	-0,420	-0,091	-0,154	0,034	0,017	-0,000	-0,019	0,012	-0,025	0,018	-0,002	-0,028
CatBoost	-0,552	-0,076	-0,042	0,025	0,015	0,000	-0,016	0,010	-0,026	0,026	0,005	0,015
MLP	-0,331	-0,099	-0,008	-0,003	-0,004	-0,002	-0,006	-0,002	-0,020	-0,004	-0,046	-0,006
ResNet	-0,680	-0,031	0,037	-0,002	-0,013	0,003	-0,000	0,016	0,008	0,000	0,007	-0,005
TabNet	0,022	0,013	0,006	0,004	0,006	-0,373	-0,084	0,001	-0,002	0,015	0,011	0,004
SAINT	-0,134	-0,001	-0,024	-0,012	-0,053	-0,011	0,036	0,003	0,051	0,034	-0,004	0,026
FastFormer	-0,269	-0,009	0,034	-0,065	-0,052	-0,005	0,012	-0,059	0,010	-0,015	-0,007	-0,017
FT-Transformer	-0,370	-0,000	-0,018	-0,030	-0,026	-0,006	0,015	-0,044	0,027	-0,029	-0,021	0,014
TabTransformer	-0,440	-0,018	-0,012	0,003	-0,112	0,001	-0,000	-0,012	0,023	-0,016	0,010	-0,022
Média	-0,233	0,006	0,005	0,015	-0,006	-0,021	-0,002	0,004	-0,003	0,010	-0,009	-0,001
Desvio Padrão	0,366	0,101	0,068	0,034	0,034	0,085	0,027	0,025	0,022	0,017	0,027	0,013

Tabela 4 – SHAP Absolutos para a Base de Dados dos Atrasos no SUS

Modelo	Diferença Datas	SMS Paciente	Dia Marcada	Dia Consulta	Idade 18-24	Idade 13-17	Idade 60-64	>2 Deficiências	Idade 65-69	Marcada 4ª-Feira	Bolsa-Família	Mulher
LASSO	0,241	0,160	0,082	0,018	0,058	0,047	0,038	0,000	0,026	0,000	0,030	0,009
Ridge	0,221	0,151	0,078	0,021	0,052	0,042	0,038	0,000	0,027	0,001	0,029	0,012
Elastic Net	0,241	0,161	0,082	0,018	0,057	0,046	0,038	0,000	0,027	0,000	0,030	0,009
KNN	0,135	0,091	0,068	0,050	0,015	0,011	0,019	0,005	0,015	0,035	0,010	0,021
SVM Linear	0,066	0,036	0,019	0,007	0,015	0,012	0,008	0,000	0,006	0,094	0,007	0,003
Decision Tree	0,606	0,077	0,004	0,003	0,006	0,005	0,022	0,001	0,015	0,021	0,023	0,004
Random Forest	0,544	0,079	0,065	0,029	0,017	0,007	0,018	0,001	0,015	0,007	0,017	0,010
Extra Trees	0,368	0,095	0,049	0,039	0,043	0,018	0,026	0,001	0,023	0,011	0,031	0,009
XGBoost	0,518	0,065	0,041	0,045	0,025	0,014	0,018	0,001	0,014	0,011	0,020	0,012
LightGBM	0,395	0,068	0,124	0,069	0,026	0,013	0,019	0,001	0,015	0,011	0,014	0,027
CatBoost	0,451	0,055	0,057	0,047	0,019	0,012	0,019	0,001	0,012	0,016	0,019	0,018
MLP	0,243	0,064	0,021	0,030	0,015	0,018	0,038	0,002	0,027	0,032	0,014	0,020
ResNet	0,446	0,031	0,051	0,053	0,027	0,027	0,025	0,004	0,015	0,011	0,016	0,016
TabNet	0,022	0,013	0,006	0,001	0,004	0,006	0,002	0,373	0,084	0,029	0,015	0,004
SAINT	0,116	0,024	0,035	0,027	0,019	0,042	0,037	0,009	0,028	0,013	0,026	0,045
FastFormer	0,233	0,026	0,031	0,024	0,053	0,043	0,015	0,006	0,015	0,009	0,022	0,015
FT-Transformer	0,245	0,032	0,033	0,026	0,037	0,029	0,030	0,006	0,021	0,030	0,026	0,018
TabTransformer	0,327	0,027	0,015	0,014	0,019	0,084	0,023	0,005	0,005	0,052	0,013	0,017
Média	0,301	0,070	0,048	0,029	0,028	0,027	0,024	0,023	0,022	0,021	0,020	0,015
Desvio Padrão	0,163	0,046	0,031	0,018	0,017	0,020	0,010	0,085	0,017	0,022	0,007	0,010

Tabela 5 – Feature Permutation para a Base de Dados dos Atrasos no SUS

Modelo	Diferença Datas	Marcada 6ª-Feira	Idade 18-24	Consulta 6ª-Feira	Dia Marcada	SMS Paciente	Mulher	Marcada 4ª-Feira	Dia Consulta	Marcada 2ª-Feira	Marcada 3ª-Feira	Bolsa-Família
LASSO	-0,161	-0,022	0,270	0,076	0,080	-0,115	0,015	0	-0,023	0	0	-0,016
Ridge	-0,233	-0,032	0,248	0,066	0,056	-0,116	0,018	-0,002	-0,023	-0,001	0,001	-0,014
Elastic Net	-0,188	-0,026	0,275	0,075	0,062	-0,106	0,014	0	-0,016	0	0	-0,015
KNN	-0,003	-0,314	0,128	-0,070	-0,039	-0,003	0,152	0,006	-0,008	0,014	0,016	-0,063
SVM Linear	-0,099	0,147	0,069	-0,183	0,011	-0,024	0,004	-0,102	-0,005	-0,036	-0,084	-0,003
Decision Tree	-0,392	0	0,149	0	0	0	0,122	-0,063	0	-0,085	-0,111	0
Random Forest	-0,309	-0,183	0,124	0,036	-0,029	-0,036	0,089	-0,053	-0,035	-0,042	-0,022	-0,001
Extra Trees	-0,179	-0,175	0,088	-0,045	-0,101	-0,119	0,057	-0,031	-0,093	-0,022	-0,017	-0,039
XGBoost	-0,348	-0,072	0,124	0,014	-0,068	-0,036	0,089	-0,037	-0,024	-0,050	-0,038	-0,002
LightGBM	-0,267	-0,196	0,097	0,003	-0,097	-0,023	0,084	-0,017	-0,057	-0,041	-0,015	-0,014
CatBoost	-0,274	-0,112	0,038	-0,381	-0,020	-0,021	0,037	-0,011	-0,011	-0,007	-0,007	-0,016
MLP	-0,115	-0,270	0,016	-0,239	-0,053	-0,040	0,028	-0,046	-0,030	-0,025	-0,018	-0,002
ResNet	-0,294	0,002	0,001	0,006	-0,273	-0,090	0,011	0,006	-0,206	0,010	0,004	-0,003
TabNet	0,115	0,006	0,026	0,014	0,169	0,084	-0,012	0,036	0,056	0,008	0,031	0,011
SAINT	-0,191	-0,057	-0,001	-0,019	-0,069	-0,115	-0,004	-0,038	-0,052	-0,052	-0,046	0,014
FastFormer	-0,178	-0,081	0,003	-0,055	-0,018	-0,081	0,001	-0,073	-0,011	-0,092	-0,033	0,000
FT-Transformer	-0,165	-0,089	0,001	-0,084	-0,032	-0,035	0,003	-0,085	0,000	-0,109	-0,046	-0,001
TabTransformer	-0,175	-0,047	0,001	-0,041	-0,030	-0,116	0,001	-0,099	-0,007	-0,043	-0,101	0,000
Média	-0,192	-0,084	0,092	-0,046	-0,025	-0,055	0,039	-0,034	-0,030	-0,032	-0,027	-0,009
Desvio Padrão	0,118	0,107	0,092	0,114	0,089	0,054	0,047	0,038	0,052	0,035	0,038	0,017

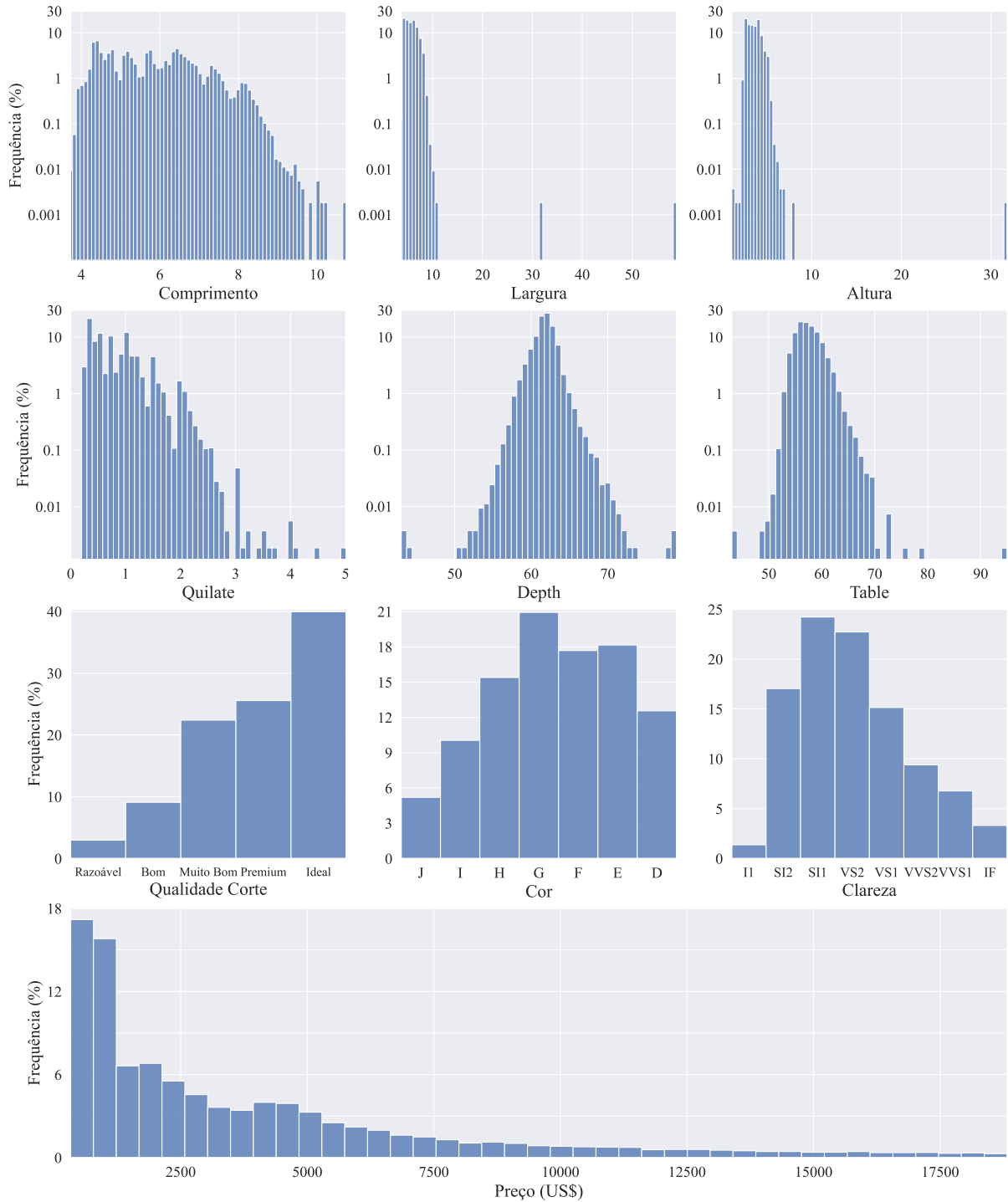
Tabela 6 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Atrasos no SUS

Modelo	Diferença Datas	SMS Paciente	Dia Marcada	Dia Consulta	Idade 13-17	Hiper-tensão	Idade 18-24	Idade 60-64	Idade 80-89	Consulta 4ª-Feira	Marcada 2ª-Feira	Bolsa-Família
LASSO	0,202	0,102	0,057	-0,013	0,062	0,000	0,057	-0,049	-0,027	-0,001	0,000	0,031
Ridge	0,184	0,096	0,053	-0,015	0,056	0,004	0,051	-0,049	-0,029	-0,011	0,002	0,029
Elastic Net	0,201	0,103	0,057	-0,013	0,062	0,000	0,057	-0,049	-0,027	-0,001	0,000	0,031
Decision Tree	0,692	0,013	0,001	0,002	0,016	0,009	0,011	0,010	0,001	0,050	0,049	0,008
Random Forest	0,573	0,044	0,041	0,041	0,018	0,009	0,012	0,010	0,002	0,010	0,032	0,008
Extra Trees	0,348	0,104	0,087	0,063	0,016	0,011	0,013	0,011	0,004	0,014	0,037	0,013
XGBoost	0,099	0,080	0,016	0,021	0,045	0,013	0,034	0,034	0,019	0,026	0,040	0,021
LightGBM	0,162	0,026	0,156	0,135	0,014	0,025	0,014	0,013	0,007	0,023	0,020	0,021
CatBoost	0,146	0,022	0,103	0,094	0,011	0,030	0,008	0,016	0,015	0,026	0,019	0,034
TabNet	0,394	0,068	0,000	0,000	0,000	0,174	0,000	0,000	0,087	0,056	0,000	0,000
Média Absolutos	0,300	0,066	0,057	0,040	0,030	0,027	0,026	0,024	0,022	0,022	0,020	0,020
Desvio Padrão dos Absolutos	0,188	0,035	0,046	0,042	0,022	0,050	0,021	0,018	0,024	0,018	0,018	0,011

As tabelas 3,4,5 e 6 retratam a alta variabilidade das importâncias das variáveis atribuídas por cada modelo, mesmo quando o resultado de suas métricas não diferiu de forma acentuada, esse resultado se repetiu para outras bases de dados em estudo, demonstrando que possuem formas diferentes de entender os problemas em questão.

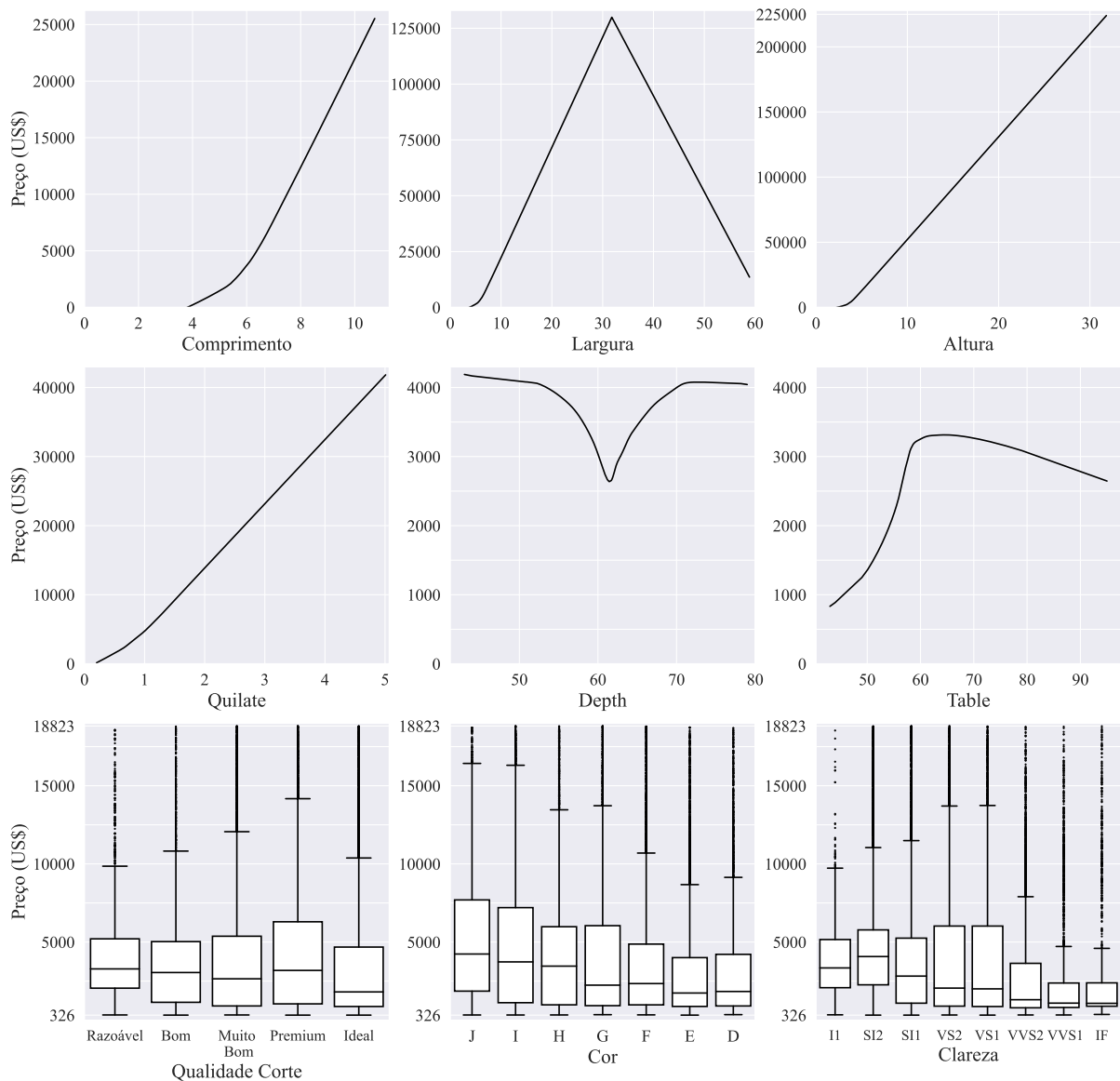
4.2 PREÇO DE DIAMANTES

Figura 27 – Histogramas das Variáveis Numéricas para a Base de Dados dos Diamantes



Fonte: Figura do Autor

Figura 28 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Diamantes



Fonte: Figura do Autor

A variável Cor é categórica e recebe os valores na ordem contrária à alfabética, indo da letra J, o de pior qualidade, até D, o de melhor qualidade.

A variável Clareza também é categórica e recebe os valores, do pior para o melhor, respectivamente, I1, SI2, SI1, VS2, VS1, VVS2, VVS1 e IF. Também há a variável que trata da qualidade do corte, indo da categoria Razoável até a Ideal.

A variável Table é a razão da largura do topo do diamante e a de seu ponto mais largo. A variável Depth é a porcentagem da razão entre a altura e a média das variáveis Largura e Comprimento.

Tabela 7 – Métricas para a Base de Dados dos Diamantes

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	0,020	0,110	1,097	0,981	0,982	0,71	45
Ridge	0,020	0,110	1,098	0,981	0,982	0,02	1
Elastic Net	0,020	0,110	1,096	0,981	0,982	0,81	52
Relaxed Lasso	0,020	0,110	1,085	0,981	0,982	1,59	102
KNN	0,017	0,096	1,029	0,983	0,984	0,31	20
SVM	0,012	0,087	0,970	0,988	0,989	90,88	5801
Decision Tree	0,023	0,112	1,084	0,978	0,979	0,25	16
Random Forest	0,021	0,108	1,035	0,980	0,981	57,66	3680
Extra Trees	0,015	0,094	1,049	0,986	0,986	35,19	2246
XGBoost	0,012	0,082	1,035	0,989	0,989	2,70	173
LightGBM	0,010	0,077	1,008	0,990	0,991	5,03	321
CatBoost	0,011	0,079	1,022	0,990	0,991	40,30	2573
MLP	0,012	0,084	1,006	0,988	0,989	67,21	4290
ResNet	0,010	0,072	1,036	0,991	0,991	267,90	17101
TabNet	0,030	0,122	3,301	0,971	0,973	213,35	13619
SAINT	0,130	0,294	1,565	0,875	0,875	83,02	5299
FastFormer	0,036	0,153	0,875	0,966	0,971	360,74	23027
FT-Transformer	0,017	0,099	1,024	0,984	0,984	107,24	6846
TabTransformer	0,017	0,103	0,834	0,984	0,984	346,54	22120
Trivial	1,037	0,878	2,030	0	0	-	-

A tabela 7 constata os melhores resultados obtidos pelo ResNet e MLP sobre as outras redes neurais, de forma similar ao percebido nas outras bases de dados em estudo, demonstrando que as suas arquiteturas menos complexas não foram impedimento à qualidade da predição.

Tabela 8 – SHAP Médio para a Base de Dados dos Diamantes

Modelo	Comprimento	Quilate	Clareza I1	Largura	Clareza SI2	Clareza IF	Clareza SI1	Cor I	Cor J	Altura	Clareza VS1	Clareza VVS1
LASSO	0,287	-0,110	-0,080	0,004	-0,068	0,112	0,049	-0,018	0,000	0,004	-0,002	-0,032
Ridge	0,264	-0,101	-0,061	0,003	-0,038	0,141	0,010	-0,006	0,000	0,004	-0,006	-0,045
Elastic Net	0,297	-0,114	-0,076	0,004	-0,056	0,138	0,030	-0,031	-0,000	0,004	-0,005	-0,042
Relaxed Lasso	0,304	-0,116	-0,083	0,004	-0,069	0,124	0,047	-0,028	-0,000	0,004	-0,003	-0,036
KNN	-0,116	0,120	-0,018	-0,106	-0,037	0,034	0,036	-0,043	-0,036	-0,108	0,038	0,045
SVM	-0,408	-0,110	-0,012	0,253	-0,020	0,012	-0,007	-0,009	-0,008	-0,056	0,007	0,010
Decision Tree	-0,152	-0,250	-0,023	0,037	-0,102	0,005	-0,125	-0,052	-0,037	0,053	-0,021	0,017
Random Forest	-0,089	-0,146	-0,024	-0,105	-0,124	0,004	-0,134	-0,061	-0,043	0,035	-0,028	0,006
Extra Trees	-0,260	-0,449	-0,029	0,012	-0,037	0,010	-0,043	-0,019	-0,005	0,049	-0,016	0,006
XGBoost	-0,210	-0,176	-0,019	0,091	-0,093	0,020	-0,062	-0,039	-0,032	0,026	0,012	0,024
LightGBM	-0,080	0,001	0,215	-0,033	0,046	-0,138	0,012	0,051	0,100	-0,010	-0,005	-0,073
CatBoost	-0,172	-0,195	-0,023	0,013	-0,100	0,029	-0,070	-0,049	-0,038	-0,008	0,020	0,036
MLP	-0,332	-0,160	0,008	-0,088	0,046	0,017	0,010	0,015	0,023	-0,192	0,010	0,014
ResNet	-0,620	-0,042	-0,013	0,114	0,005	0,003	0,001	-0,014	-0,013	-0,038	-0,005	-0,017
TabNet	0,038	0,002	-0,336	0,057	-0,054	0,054	0,026	0,012	-0,199	0,006	-0,002	0,050
SAINT	-0,019	-0,002	0,116	0,005	0,044	0,051	-0,002	0,051	0,056	-0,018	0,036	0,086
FastFormer	0,077	0,092	0,054	-0,219	0,025	-0,005	-0,011	0,152	0,008	-0,042	0,005	0,012
FT-Transformer	-0,053	-0,018	0,149	-0,020	0,051	0,009	0,042	0,092	0,080	-0,017	-0,033	-0,007
TabTransformer	0,007	0,005	0,076	0,005	0,062	-0,016	-0,054	-0,006	0,068	0,005	0,413	-0,031
Média	-0,065	-0,093	-0,009	0,002	-0,027	0,032	-0,013	-0,000	-0,004	-0,016	0,022	0,001
Desvio Padrão	0,242	0,126	0,109	0,094	0,058	0,063	0,053	0,052	0,061	0,055	0,094	0,038

Tabela 9 – SHAP Absolutos para a Base de Dados dos Diamantes

Modelo	Comprimento	Quilate	Largura	Altura	Clareza I1	Clareza SI2	Cor J	Clareza VS1	Clareza SI1	Cor I	Cor D	Clareza VVS1
LASSO	0,528	0,159	0,007	0,013	0,009	0,038	0,011	0,006	0,025	0,011	0,021	0,013
Ridge	0,503	0,151	0,007	0,012	0,007	0,023	0,007	0,018	0,005	0,004	0,027	0,018
Elastic Net	0,556	0,167	0,007	0,013	0,008	0,032	0,016	0,013	0,016	0,020	0,012	0,017
Relaxed Lasso	0,562	0,167	0,008	0,014	0,009	0,039	0,015	0,008	0,024	0,018	0,015	0,015
KNN	0,186	0,127	0,178	0,181	0,011	0,023	0,017	0,019	0,017	0,018	0,015	0,019
SVM	0,227	0,320	0,096	0,043	0,007	0,038	0,017	0,014	0,022	0,020	0,020	0,019
Decision Tree	0,067	0,196	0,460	0,024	0,011	0,044	0,017	0,012	0,054	0,023	0,016	0,009
Random Forest	0,056	0,175	0,489	0,015	0,010	0,050	0,018	0,012	0,054	0,025	0,005	0,003
Extra Trees	0,304	0,277	0,101	0,049	0,010	0,068	0,022	0,008	0,055	0,023	0,010	0,005
XGBoost	0,198	0,269	0,230	0,081	0,008	0,039	0,014	0,005	0,026	0,016	0,015	0,011
LightGBM	0,112	0,094	0,116	0,052	0,147	0,032	0,068	0,004	0,010	0,035	0,074	0,050
CatBoost	0,173	0,297	0,165	0,143	0,008	0,035	0,013	0,007	0,025	0,017	0,015	0,013
MLP	0,190	0,240	0,120	0,107	0,014	0,046	0,023	0,022	0,018	0,024	0,023	0,026
ResNet	0,249	0,344	0,087	0,017	0,010	0,028	0,014	0,018	0,017	0,018	0,022	0,025
TabNet	0,038	0,002	0,057	0,006	0,336	0,054	0,199	0,002	0,026	0,012	0,023	0,050
SAINT	0,086	0,088	0,101	0,107	0,063	0,024	0,036	0,029	0,021	0,031	0,013	0,047
FastFormer	0,149	0,066	0,137	0,147	0,039	0,056	0,021	0,003	0,030	0,087	0,049	0,010
FT-Transformer	0,081	0,104	0,125	0,082	0,091	0,058	0,055	0,022	0,040	0,059	0,036	0,013
TabTransformer	0,039	0,030	0,023	0,027	0,065	0,063	0,058	0,347	0,050	0,008	0,008	0,028
Média	0,227	0,172	0,132	0,060	0,045	0,042	0,034	0,030	0,028	0,025	0,022	0,021
Desvio Padrão	0,176	0,096	0,132	0,053	0,078	0,013	0,043	0,075	0,015	0,019	0,016	0,014

Tabela 10 – Feature Permutation para a Base de Dados dos Diamantes

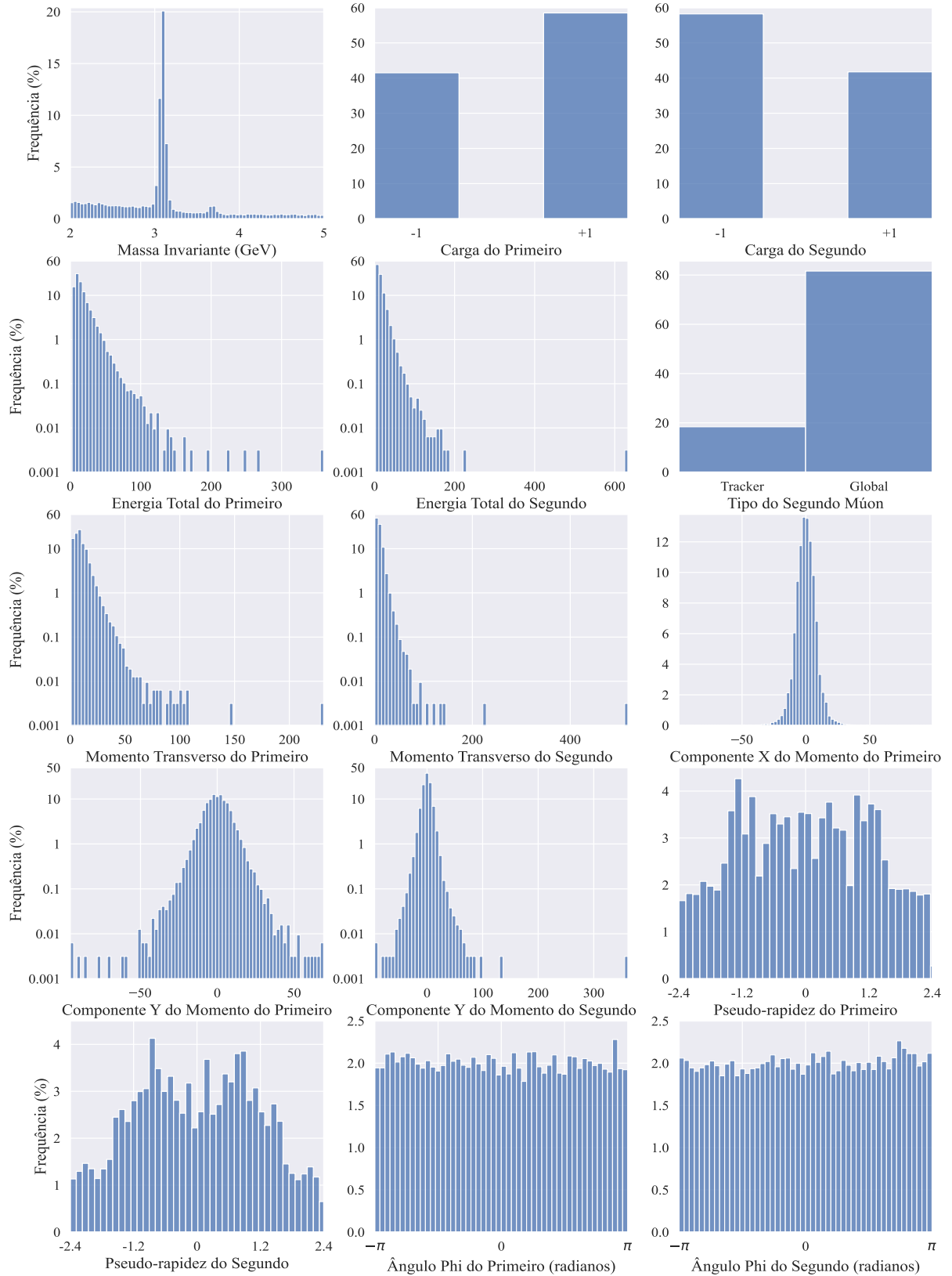
Modelo	Clareza I1	Comprimento	Largura	Quilate	Altura	Cor E	Clareza SI2	Clareza SI1	Clareza VS2	Clareza VS1	Qualidade Corte Bom	Cor H
LASSO	-0,168	0,496	0,007	-0,173	0,013	0,030	0,008	0,008	0,001	-0,002	0,011	0,001
Ridge	-0,166	0,456	0,006	-0,167	0,013	0,053	0,009	0,003	-0,006	-0,007	0,014	-0,003
Elastic Net	-0,201	0,469	0,007	-0,176	0,013	0,017	0,013	0,006	-0,005	-0,003	-0,005	0,004
Relaxed Lasso	-0,187	0,493	0,007	-0,179	0,014	0,021	0,015	0,011	0,000	-0,002	-0,005	0,003
KNN	-0,447	0,066	0,066	0,081	0,044	0,120	0,003	0,003	0,003	0,001	0,031	0,002
SVM	-0,397	0,071	0,129	0,108	0,105	-0,003	0,010	0,012	0,010	0,010	-0,030	0,004
Decision Tree	-0,480	0	0	0	0	0	-0,050	0	0	0	0	0
Random Forest	-0,481	0,009	0,424	0,039	0,011	0	-0,035	-0,000	0	0	0	0,000
Extra Trees	-0,494	0,195	0,076	0,156	0,040	0,006	-0,020	-0,005	-0,001	-0,000	0,000	0,000
XGBoost	-0,309	0,151	0,228	0,142	0,071	0,020	-0,000	0,007	-0,005	-0,012	-0,001	0,004
LightGBM	-0,357	0,144	0,117	0,137	0,090	0,040	-0,005	0,001	-0,009	-0,015	-0,010	0,001
CatBoost	-0,307	0,126	0,140	0,194	0,113	0,019	0,011	-0,002	-0,006	-0,011	-0,014	0,003
MLP	-0,467	0,102	0,068	0,182	0,058	0,018	0,013	0,010	-0,005	-0,010	-0,010	0,002
ResNet	0,001	0,277	0,412	0,135	0,009	0,021	0,007	0,003	0,017	0,008	0,000	0,009
TabNet	-0,017	0,121	0,039	0,027	0,030	0,015	-0,047	0,007	0,000	-0,014	0,074	0,143
SAINT	-0,002	0,149	0,252	0,218	0,189	0,010	-0,022	-0,008	0,001	0,005	-0,002	0,002
FastFormer	-0,029	-0,022	-0,450	0,077	-0,181	-0,026	-0,014	-0,002	-0,011	0,004	0,023	0,006
FT-Transformer	0,000	-0,092	-0,118	-0,124	-0,116	0,040	-0,016	0,036	-0,026	0,011	0,009	0,043
TabTransformer	0,003	-0,014	-0,012	-0,010	-0,012	0,013	0,136	0,244	0,241	0,184	0,003	0,011
Média	-0,237	0,168	0,097	0,035	0,027	0,022	0,001	0,018	0,010	0,008	0,005	0,012
Desvio Padrão	0,187	0,181	0,201	0,134	0,078	0,029	0,038	0,054	0,055	0,042	0,021	0,032

Tabela 11 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Diamantes

Modelo	Comprimento	Largura	Quilate	Altura	Clareza SI2	Clareza SI2	Clareza VVS2	Cor J	Clareza I1	Cor D	Clareza VVS1	Cor I
LASSO	0,482	0,006	-0,152	0,012	-0,040	-0,023	0,017	-0,020	-0,028	0,025	0,021	-0,014
Ridge	0,456	0,006	-0,144	0,011	-0,023	-0,005	0,027	-0,013	-0,022	0,032	0,029	-0,005
Elastic Net	0,499	0,007	-0,157	0,012	-0,033	-0,014	0,024	-0,028	-0,027	0,015	0,027	-0,024
Relaxed Lasso	0,506	0,007	-0,158	0,012	-0,040	-0,021	0,019	-0,027	-0,029	0,018	0,023	-0,022
Decision Tree	0,009	0,870	0,075	0,000	0,010	0,007	0,001	0,005	0,006	0,001	0,001	0,004
Random Forest	0,024	0,799	0,133	0,000	0,010	0,007	0,001	0,005	0,006	0,001	0,001	0,004
Extra Trees	0,471	0,092	0,327	0,060	0,015	0,008	0,001	0,004	0,006	0,001	0,002	0,004
XGBoost	0,254	0,288	0,185	0,038	0,036	0,025	0,021	0,021	0,027	0,008	0,017	0,016
LightGBM	0,021	0,020	0,014	0,012	0,058	0,059	0,037	0,033	0,022	0,045	0,028	0,042
CatBoost	0,153	0,173	0,234	0,239	0,042	0,018	0,012	0,020	0,019	0,008	0,013	0,015
TabNet	0,095	0,143	0,009	0,018	0,052	0,124	0,066	0,041	0,023	0,031	0,009	0,016
Média Absolutos	0,270	0,219	0,145	0,038	0,033	0,028	0,021	0,020	0,020	0,017	0,016	0,015
Desvio Padrão dos Absolutos	0,205	0,303	0,087	0,066	0,015	0,033	0,018	0,012	0,009	0,014	0,010	0,011

4.3 MASSA INVARIANTE NA COLISÃO DE MUÓNS

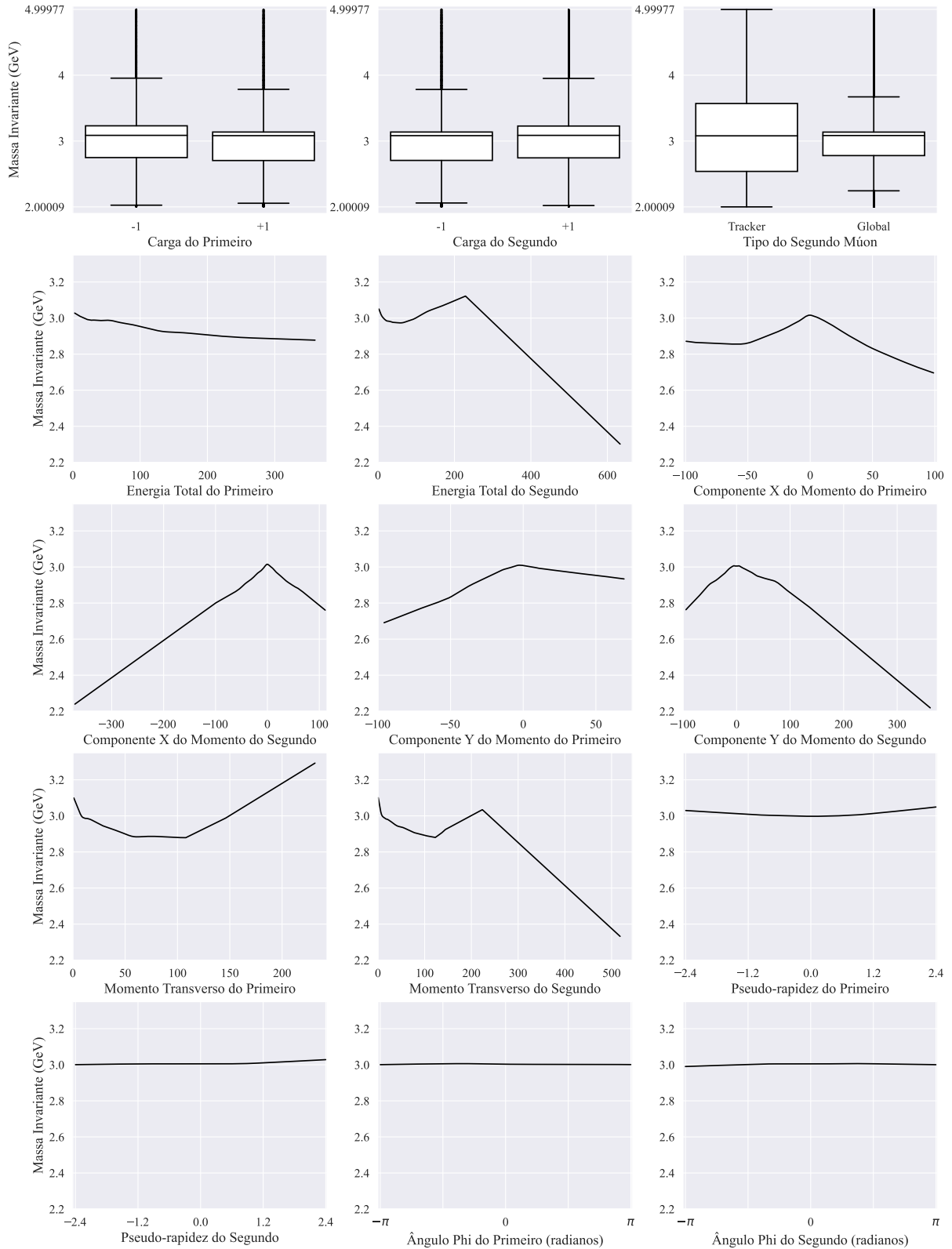
Figura 29 – Histogramas das Variáveis para a Base de Dados dos Muóns



Fonte: Figura do Autor

A variável 2º Múon Global é binária e descreve o tipo do segundo múon envolvido na colisão: 1 representa um múon global e 0 um múon *tracker*.

Figura 30 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Múons



Fonte: Figura do Autor

Tabela 12 – Métricas para a Base de Dados dos Múons

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	0,387	0,425	2,225	0,020	0,020	0,03	5
Ridge	0,386	0,425	2,183	0,021	0,021	0,01	1
Elastic Net	0,386	0,425	2,225	0,020	0,020	0,04	6
Relaxed Lasso	0,387	0,425	2,225	0,020	0,020	0,03	5
KNN	0,349	0,427	2,200	0,116	0,123	0,09	15
SVM	0,283	0,383	2,655	0,283	0,293	44,13	7092
Decision Tree	0,378	0,439	2,087	0,043	0,043	0,97	155
Random Forest	0,343	0,418	2,012	0,131	0,131	15,88	2552
Extra Trees	0,361	0,426	2,001	0,084	0,084	9,97	1603
XGBoost	0,251	0,374	2,252	0,362	0,364	3,75	603
LightGBM	0,153	0,290	2,269	0,613	0,615	2,26	363
CatBoost	0,123	0,257	1,938	0,688	0,691	6,53	1049
MLP	0,046	0,147	2,003	0,884	0,887	269,38	43287
ResNet	0,033	0,124	2,027	0,917	0,919	257,88	41440
TabNet	0,342	0,424	2,208	0,133	0,133	37,07	5957
SAINT	0,140	0,279	2,063	0,645	0,676	416,40	66914
FastFormer	0,066	0,167	2,333	0,833	0,834	303,07	48702
FT-Transformer	0,063	0,176	2,070	0,840	0,843	307,93	49483
TabTransformer	0,362	0,435	2,035	0,082	0,085	202,11	32478
Trivial	0,394	0,418	1,914	0	0	-	-

Tabela 13 – SHAP Médio para a Base de Dados dos Múons

Modelo	Pseudo-rapidez do 1º	Ângulo ϕ do 2º	Transverse Momentum do 1º	Energia Total do 2º	Pseudo-rapidez do 2º	2º Muón Global	Carga do 1º	Transverse Momentum do 2º	Componente Y do Momento do 2º	Carga do 2º
LASSO	0	-0,233	0,121	-0,246	0	0,108	0,112	0,042	0	0,117
Ridge	-0,144	0,010	0,017	0,005	-0,051	0,019	0,020	0,014	0,176	0,020
Elastic Net	0	-0,500	0,088	0,008	0	0,101	0,092	0,055	0	0,093
Relaxed Lasso	0	0	0,089	-0,429	0	0,126	0,125	0,033	0	0,127
KNN	0,141	0,110	0,089	0,013	0,029	0,027	0,085	0,074	0,093	0,086
SVM	-0,038	-0,055	-0,029	-0,054	-0,058	-0,022	-0,031	-0,034	-0,104	-0,031
Decision Tree	0,168	0,002	0,295	0,167	-0,028	0,025	0,000	0,161	0,016	0,000
Random Forest	0,123	-0,010	0,147	0,100	-0,077	-0,015	0,078	0,129	0,001	0,022
Extra Trees	0,121	-0,006	0,067	0,037	-0,045	-0,052	-0,249	0,069	-0,007	-0,268
XGBoost	0,085	-0,052	0,142	0,044	-0,148	0,001	0,028	-0,058	0,079	-0,003
LightGBM	0,147	-0,052	0,015	-0,004	-0,074	0,055	-0,001	-0,046	0,146	0,005
CatBoost	0,157	-0,106	-0,030	-0,038	-0,117	-0,004	0,005	-0,093	0,101	0,003
MLP	0,188	0,207	-0,052	-0,021	0,120	0,003	0,000	-0,056	0,012	0,001
ResNet	0,187	0,215	-0,001	-0,004	0,291	0,000	-0,001	0,046	0,001	-0,000
TabNet	0,087	0,106	0,081	-0,029	-0,030	-0,173	0,063	-0,012	0,017	0,044
SAINT	0,081	0,087	-0,125	-0,082	-0,080	0,006	-0,001	-0,033	0,120	0,003
FastFormer	-0,164	-0,021	-0,032	-0,039	-0,070	0,082	0,072	-0,059	0,058	0,023
FT-Transformer	0,242	0,007	0,030	-0,036	0,099	0,010	0,002	-0,004	0,097	0,002
TabTransformer	0,002	0,028	0,031	0,033	-0,009	0,426	-0,214	0,030	0,011	-0,146
Média	0,073	-0,014	0,050	-0,030	-0,013	0,038	0,010	0,014	0,043	0,005
Desvio Padrão	0,108	0,153	0,090	0,122	0,096	0,112	0,094	0,065	0,065	0,087

Tabela 14 – SHAP Absolutos para a Base de Dados dos Múons

Modelo	Transverse Momentum do 1º	Transverse Momentum do 2º	Pseudo-rapidez do 1º	Pseudo-rapidez do 2º	Carga do 1º	Ângulo ϕ do 2º	Energia Total do 2º	2º Muón Global	Ângulo ϕ do 1º	Energia Total do 1º
LASSO	0,340	0,225	0	0	0,148	0,000	0,067	0,074	0	0,086
Ridge	0,244	0,148	0,007	0,019	0,112	0,042	0,085	0,071	0,060	0,068
Elastic Net	0,340	0,213	0	0	0,145	0,002	0,074	0,073	0	0,093
Relaxed Lasso	0,340	0,231	0	0	0,149	0,000	0,064	0,075	0	0,083
KNN	0,066	0,054	0,109	0,052	0,072	0,091	0,033	0,041	0,089	0,028
SVM	0,044	0,080	0,035	0,051	0,026	0,046	0,080	0,018	0,041	0,109
Decision Tree	0,284	0,161	0,162	0,028	0,001	0,002	0,188	0,024	0,019	0,050
Random Forest	0,130	0,098	0,094	0,066	0,052	0,047	0,094	0,017	0,039	0,041
Extra Trees	0,076	0,063	0,110	0,050	0,247	0,011	0,046	0,074	0,007	0,020
XGBoost	0,102	0,061	0,091	0,109	0,017	0,053	0,052	0,006	0,060	0,070
LightGBM	0,043	0,064	0,119	0,117	0,009	0,071	0,034	0,031	0,068	0,056
CatBoost	0,025	0,073	0,129	0,142	0,003	0,100	0,027	0,003	0,068	0,066
MLP	0,066	0,081	0,146	0,121	0,002	0,142	0,039	0,005	0,149	0,022
ResNet	0,038	0,047	0,201	0,194	0,011	0,183	0,020	0,012	0,184	0,013
TabNet	0,081	0,012	0,087	0,030	0,063	0,106	0,029	0,173	0,012	0,023
SAINT	0,068	0,045	0,081	0,085	0,015	0,098	0,045	0,017	0,069	0,037
FastFormer	0,039	0,047	0,131	0,114	0,044	0,049	0,040	0,048	0,055	0,034
FT-Transformer	0,031	0,026	0,162	0,113	0,016	0,083	0,054	0,017	0,081	0,074
TabTransformer	0,050	0,038	0,016	0,015	0,162	0,050	0,059	0,303	0,051	0,023
Média	0,127	0,093	0,088	0,069	0,068	0,062	0,059	0,057	0,055	0,052
Desvio Padrão	0,114	0,067	0,061	0,054	0,070	0,049	0,036	0,070	0,048	0,028

Tabela 15 – Feature Permutation para a Base de Dados dos Múons

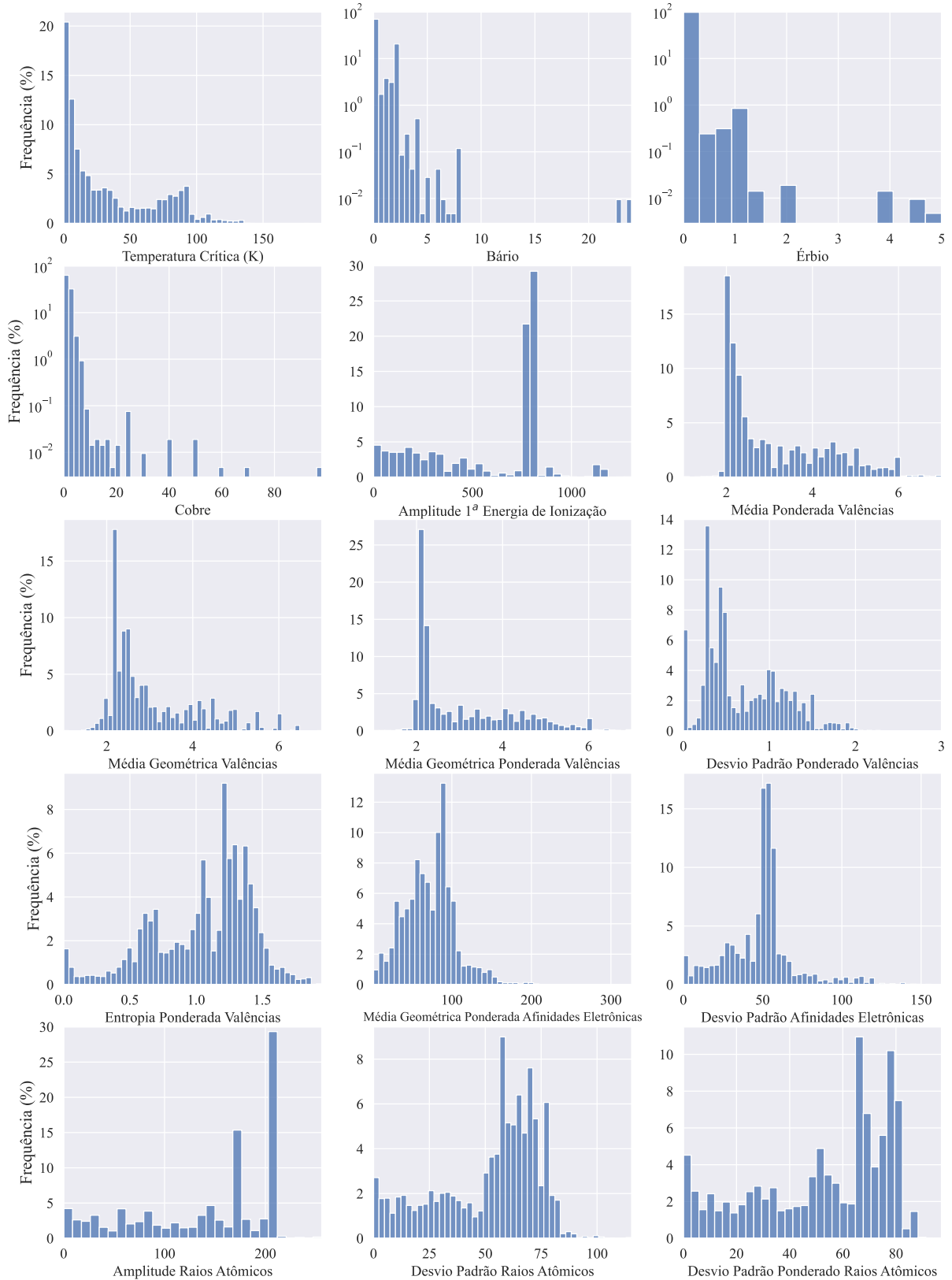
Modelo	Transverse Momentum do 1º	Pseudo-rapidez do 1º	Pseudo-rapidez do 2º	Ângulo ϕ do 2º	2º Muón Global	Transverse Momentum do 2º	Ângulo ϕ do 1º	Componente Y do Momento do 1º	Componente Y do Momento do 2º	Componente X do Momento do 1º
LASSO	-0,517	0	0	0,000	0,120	0,126	0	0	0	0
Ridge	-0,410	0,005	0,015	0,017	0,138	0,100	-0,012	0,008	0,010	-0,008
Elastic Net	-0,515	0	0	0,001	0,130	0,113	0	0	0	0
Relaxed Lasso	-0,525	0	0	0	0,126	0,122	0	0	0	0
KNN	0,091	-0,016	0,021	0,004	0,038	0,096	0,058	0,013	-0,007	0,187
SVM	0,068	-0,069	-0,097	-0,055	-0,032	-0,029	-0,018	-0,089	-0,081	-0,102
Decision Tree	-0,441	-0,142	0	0	0	-0,235	0	0	0	0
Random Forest	-0,111	-0,070	-0,065	-0,070	0,063	-0,134	-0,065	-0,139	-0,099	0,004
Extra Trees	-0,044	-0,044	-0,026	0,003	-0,464	-0,001	0,008	-0,015	-0,003	0,003
XGBoost	-0,051	-0,136	-0,074	-0,096	0,028	-0,057	-0,041	-0,095	-0,079	-0,046
LightGBM	-0,009	-0,142	-0,100	-0,118	0,011	-0,032	-0,093	-0,118	-0,079	-0,069
CatBoost	0,015	-0,204	-0,150	-0,095	-0,001	-0,007	-0,064	-0,110	-0,079	-0,087
MLP	0,145	-0,099	-0,126	-0,155	0,002	0,001	-0,109	-0,038	0,069	0,139
ResNet	-0,035	-0,247	-0,286	-0,159	0,000	-0,024	-0,155	0,006	0,009	0,001
TabNet	0,022	-0,091	-0,094	-0,136	-0,033	-0,006	-0,183	-0,180	-0,106	0,006
SAINT	-0,027	-0,135	-0,135	-0,075	-0,000	-0,033	-0,063	-0,092	-0,088	-0,099
FastFormer	-0,029	-0,115	-0,171	-0,040	-0,000	-0,020	-0,038	-0,146	-0,143	-0,126
FT-Transformer	0,002	-0,279	-0,234	-0,128	-0,000	0,001	-0,127	-0,018	-0,036	-0,028
TabTransformer	0,009	-0,014	-0,021	-0,124	-0,044	0,022	-0,118	-0,032	-0,036	-0,003
Média	-0,124	-0,095	-0,081	-0,065	0,004	0,000	-0,054	-0,055	-0,039	-0,012
Desvio Padrão	0,221	0,083	0,085	0,060	0,124	0,087	0,062	0,061	0,053	0,074

Tabela 16 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Múons

Modelo	Transverse Momentum do 1º	Transverse Momentum do 2º	Pseudo-rapidez do 1º	Energia Total do 2º	2º Muón Global	Energia Total do 1º	Carga do 1º	Pseudo-rapidez do 2º	Carga do 2º	Ângulo ϕ do 1º
LASSO	-0,354	-0,251	-0,000	-0,075	-0,070	0,096	-0,110	-0,000	0,045	-0,000
Ridge	-0,259	-0,168	-0,007	-0,097	-0,068	0,077	-0,085	-0,017	0,054	-0,052
Elastic Net	-0,354	-0,237	-0,000	-0,082	-0,069	0,103	-0,108	-0,000	0,045	-0,000
Relaxed Lasso	-0,354	-0,257	0	-0,071	-0,070	0,092	-0,111	0	0,045	0
Decision Tree	0,392	0,336	0,059	0,157	0,001	0,006	0,000	0,036	0,000	0,001
Random Forest	0,098	0,102	0,076	0,072	0,015	0,061	0,013	0,068	0,014	0,060
Extra Trees	0,106	0,083	0,178	0,048	0,160	0,027	0,048	0,095	0,158	0,011
XGBoost	0,071	0,047	0,067	0,055	0,054	0,050	0,071	0,065	0,052	0,067
LightGBM	0,071	0,073	0,079	0,064	0,004	0,066	0,004	0,081	0,005	0,075
CatBoost	0,043	0,053	0,111	0,022	0,013	0,031	0,009	0,118	0,009	0,097
TabNet	0,063	0,088	0,292	0,008	0,095	0,011	0,060	0,090	0,057	0,033
Média Absolutos	0,197	0,154	0,079	0,068	0,056	0,056	0,056	0,052	0,044	0,036
Desvio Padrão dos Absolutos	0,137	0,096	0,086	0,037	0,045	0,033	0,042	0,041	0,041	0,034

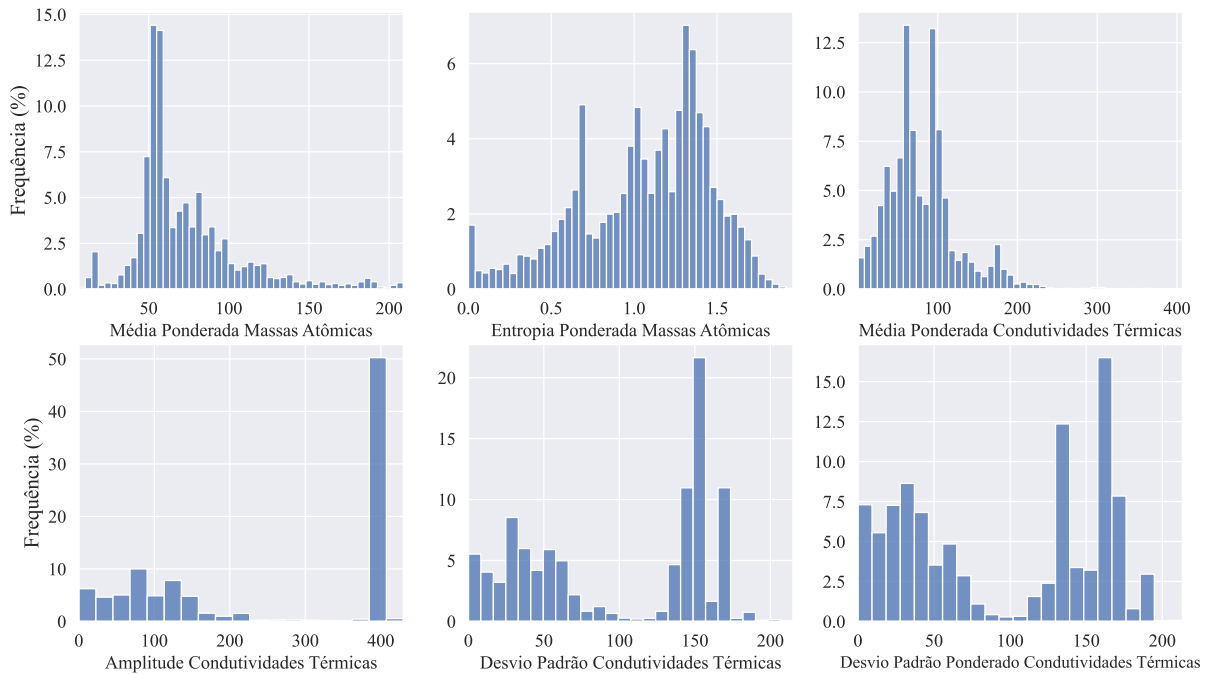
4.4 TEMPERATURA CRÍTICA DE SUPERCONDUTORES

Figura 31 – Histogramas das Variáveis para a Base de Dados dos Supercondutores



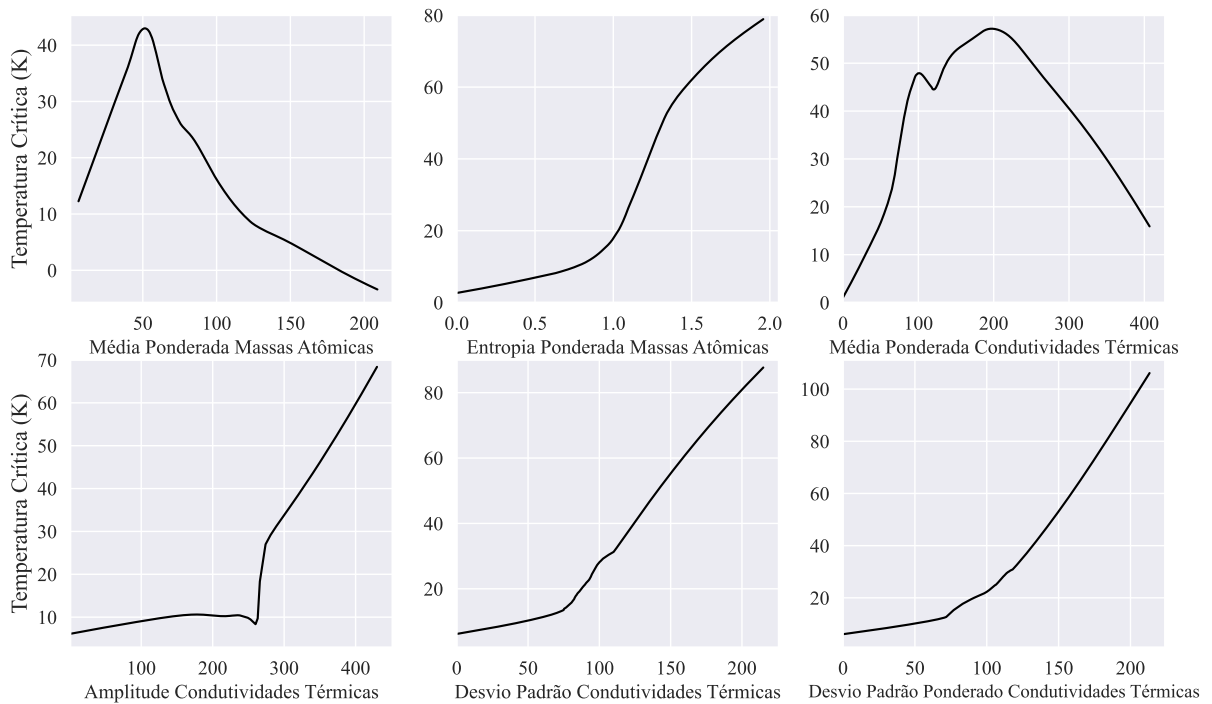
Fonte: Figura do Autor

Figura 32 – Histogramas das Variáveis para a Base de Dados dos Supercondutores



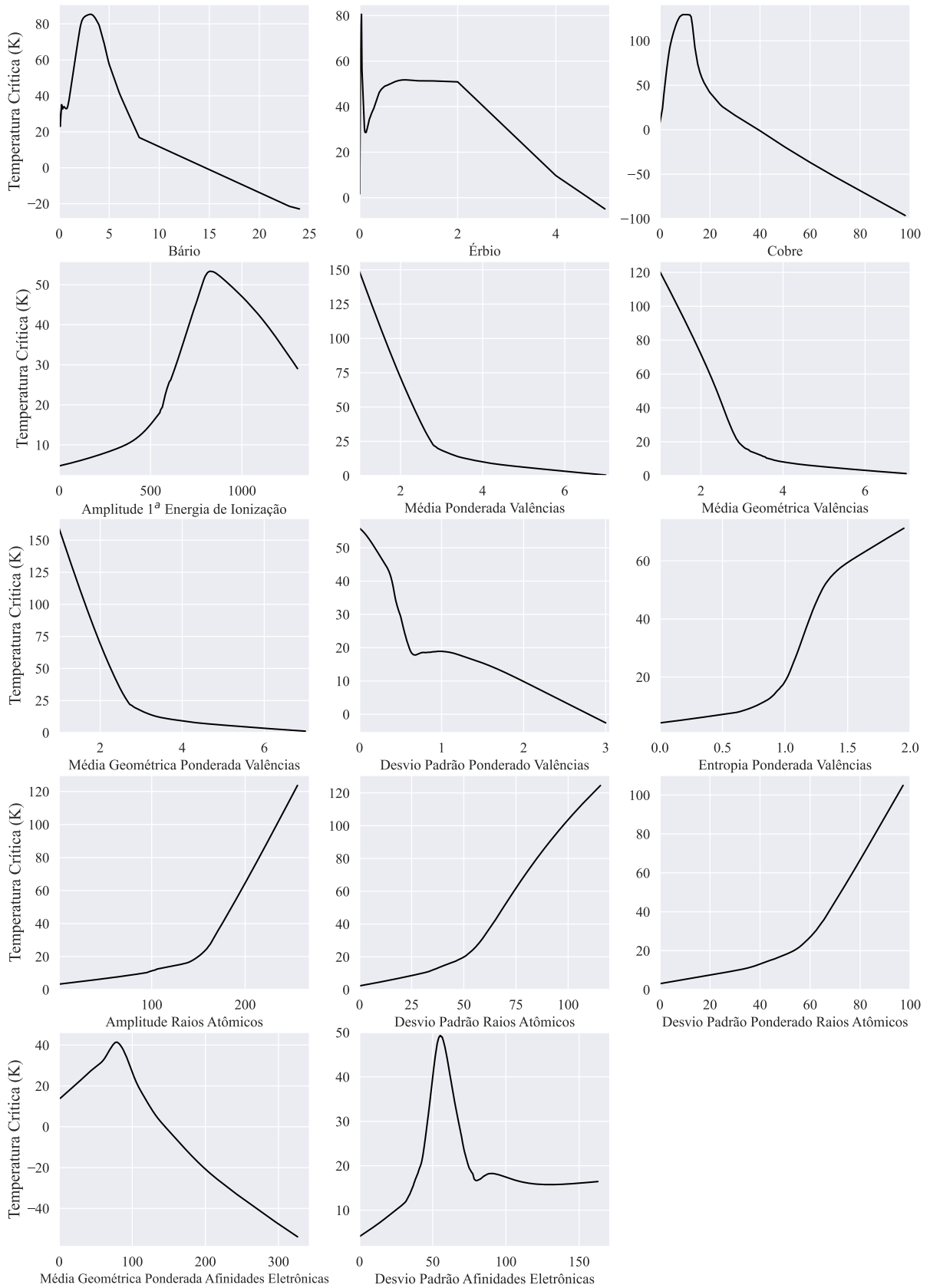
Fonte: Figura do Autor

Figura 33 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Supercondutores



Fonte: Figura do Autor

Figura 34 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Supercondutores



Fonte: Figura do Autor

O alto número de variáveis explicativas e consequente alto número de *embeddings* criados ocasionou uma excessiva necessidade de memória que impediu que o SAINT pudesse ser ajustado.

Devido a variável resposta ser contínua, positiva e com acentuada assimetria positiva, também foi ajustado um Modelo Linear Generalizado com função de ligação Gama e penalização Ridge.

Há um grande número de amostras com temperaturas críticas próximas do 0 Kelvin, como não é verossímil que seja atingida uma temperatura inferior ao zero absoluto, para todas as predições foi tomada a função de máximo entre o valor predito por cada modelo e 0,

Tabela 17 – Métricas para a Base de Dados dos Supercondutores

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	269,488	12,028	71,856	0,768	0,769	2,38	35,6
Ridge	268,181	11,975	71,173	0,769	0,770	0,07	1
Gama	378,026	13,304	256,180	0,675	0,678	0,17	2,5
Elastic Net	269,488	12,028	71,856	0,768	0,769	2,24	33,4
Relaxed Lasso	269,488	12,028	71,856	0,768	0,769	4,63	69,1
KNN	104,648	5,577	85,839	0,910	0,910	0,51	7,6
SVM	184,360	8,295	90,000	0,841	0,843	27,01	403,1
Decision Tree	200,471	8,822	75,198	0,828	0,828	0,15	2,2
Random Forest	146,723	7,909	68,109	0,874	0,875	218,57	3262,1
Extra Trees	143,253	7,901	61,278	0,877	0,877	64,35	960,4
XGBoost	116,526	7,134	61,390	0,900	0,900	1,97	29,4
LightGBM	100,998	6,336	62,074	0,913	0,915	22,56	336,7
CatBoost	103,982	6,582	69,534	0,911	0,911	262,49	3917,6
MLP	192,127	8,352	78,461	0,835	0,835	520,68	7771,0
ResNet	130,452	7,371	60,578	0,888	0,889	362,22	5405,9
TabNet	163,432	8,358	77,116	0,859	0,860	218,83	3266,0
FastFormer	127,435	7,746	67,429	0,890	0,892	686,01	10238,4
FT-Transformer	89,204	5,778	68,022	0,923	0,923	129,03	1925,7
TabTransformer	185,626	9,104	74,533	0,840	0,841	322,44	4812,3
Trivial	1162,570	29,236	101,510	0	0	-	-

É interessante notar que o KNN, mesmo enfrentando a alta dimensionalidade da base de dados, conseguiu atingir o melhor erro absoluto médio dentre os modelos estudados.

Tabela 18 – SHAP Médio para a Base de Dados dos Supercondutores

Modelo	Bário	Desvio Padrão Ponderado Condutividades Térmicas	Amplitude Condutividades Térmicas	Média Geométrica Valências	Desvio Padrão Ponderado Valências	Cobre	Entropia Ponderada Valências	Desvio Padrão Condutividades Térmicas	Amplitude Raios Atômicos	Média Geométrica Densidades	Média Geométrica Ponderada Valências
LASSO	0,052	0,004	-0,024	0,002	0,062	0,001	0,035	0,027	-0,001	-0,002	-0,000
Ridge	0,041	0,001	-0,020	-0,001	0,046	0,001	0,035	0,025	-0,000	0,003	0,007
Gama	0,045	0,056	0,039	0,011	0,014	0,027	0,014	0,035	0,035	0,029	0,009
Elastic Net	0,052	0,004	-0,024	0,002	0,062	0,001	0,035	0,027	-0,001	-0,002	-0,000
Relaxed Lasso	0,052	0,004	-0,024	0,002	0,062	0,001	0,035	0,027	-0,001	-0,002	-0,000
KNN	0,047	0,013	-0,002	0,011	0,015	0,011	0,025	0,000	0,033	0,010	0,018
SVM	0,033	0,023	0,013	0,005	0,023	-0,023	-0,028	0,031	0,006	0,011	0,008
Decision Tree	-0,230	-0,269	-0,000	0,000	-0,000	0,018	-0,004	-0,000	0,000	0,065	-0,000
Random Forest	-0,003	0,001	0,022	0,295	0,022	-0,145	0,004	-0,001	0,011	0,043	0,020
Extra Trees	0,021	0,079	0,210	0,016	0,014	0,002	0,040	0,049	0,090	0,024	0,040
XGBoost	0,023	0,056	0,047	0,006	0,002	0,004	0,022	0,010	0,028	0,004	0,066
LightGBM	-0,047	0,006	0,020	0,009	0,015	-0,028	0,007	0,019	0,030	0,041	0,054
CatBoost	0,031	0,032	0,015	0,037	0,011	0,059	0,025	0,021	0,038	0,022	0,039
MLP	0,005	0,017	0,022	0,011	-0,003	-0,021	0,011	0,020	0,006	0,018	0,025
ResNet	0,065	0,048	0,030	0,021	0,027	-0,013	-0,032	0,031	0,041	0,025	0,021
TabNet	0,003	-0,002	0,001	0,003	-0,003	-0,012	0,004	0,006	-0,000	0,004	-0,001
FastFormer	0,006	-0,004	0,007	0,007	0,026	-0,006	-0,004	0,007	0,008	0,009	-0,004
FT-Transformer	0,012	-0,001	0,044	0,023	0,019	0,006	0,001	0,009	0,005	0,009	0,012
TabTransformer	-0,001	-0,001	0,000	-0,006	-0,001	0,002	-0,004	-0,000	-0,009	-0,008	-0,007
Média	0,011	0,004	0,020	0,024	0,022	-0,006	0,012	0,018	0,017	0,016	0,016
Desvio Padrão	0,063	0,068	0,050	0,065	0,021	0,038	0,020	0,014	0,023	0,018	0,020

Tabela 19 – SHAP Absolutos para a Base de Dados dos Supercondutores

Modelo	Bário	Desvio Padrão Ponderado Condutividades Térmicas	Amplitude Condutividades Térmicas	Cobre	Média Geométrica Valências	Desvio Padrão Condutividades Térmicas	Entropia Ponderada Valências	Amplitude 1ª Energia Ionização	Desvio Padrão Afinidades Eletrônicas	Média Ponderada Condutividades Térmicas	Desvio Padrão Ponderado Valências
LASSO	0,012	0,005	0,032	0,001	0,009	0,032	0,034	0,024	0,035	0,035	0,015
Ridge	0,009	0,002	0,024	0,001	0,005	0,026	0,032	0,020	0,025	0,026	0,010
Gama	0,045	0,042	0,030	0,020	0,007	0,028	0,010	0,016	0,003	0,015	0,024
Elastic Net	0,012	0,005	0,032	0,001	0,009	0,032	0,034	0,024	0,035	0,035	0,015
Relaxed Lasso	0,012	0,005	0,032	0,001	0,009	0,032	0,034	0,024	0,035	0,035	0,015
KNN	0,016	0,012	0,009	0,011	0,006	0,009	0,011	0,008	0,006	0,013	0,012
SVM	0,047	0,022	0,005	0,016	0,002	0,019	0,021	0,011	0,015	0,023	0,031
Decision Tree	0,148	0,174	0,002	0,014	0,003	0,002	0,005	0,002	0,004	0,004	0,005
Random Forest	0,022	0,003	0,019	0,137	0,253	0,001	0,004	0,010	0,016	0,005	0,019
Extra Trees	0,079	0,079	0,160	0,022	0,010	0,040	0,024	0,046	0,002	0,005	0,009
XGBoost	0,023	0,037	0,030	0,016	0,005	0,008	0,014	0,048	0,014	0,013	0,005
LightGBM	0,050	0,006	0,015	0,058	0,006	0,015	0,006	0,011	0,008	0,010	0,010
CatBoost	0,023	0,023	0,012	0,046	0,026	0,016	0,018	0,006	0,022	0,008	0,012
MLP	0,019	0,013	0,016	0,011	0,005	0,013	0,009	0,008	0,007	0,007	0,014
ResNet	0,038	0,028	0,016	0,009	0,010	0,016	0,015	0,008	0,010	0,010	0,021
TabNet	0,003	0,002	0,001	0,012	0,003	0,006	0,004	0,001	0,015	0,002	0,003
FastFormer	0,023	0,005	0,007	0,011	0,008	0,008	0,005	0,006	0,006	0,005	0,013
FT-Transformer	0,029	0,017	0,019	0,030	0,010	0,011	0,003	0,006	0,012	0,013	0,025
TabTransformer	0,005	0,006	0,005	0,005	0,006	0,005	0,006	0,005	0,007	0,005	0,006
Média	0,033	0,025	0,024	0,022	0,021	0,017	0,015	0,015	0,015	0,014	0,014
Desvio Padrão	0,033	0,040	0,034	0,031	0,055	0,011	0,011	0,013	0,011	0,011	0,007

Tabela 20 – Feature Permutation para a Base de Dados dos Supercondutores

Modelo	Bário	Desvio Padrão Ponderado Condutividades Térmicas	Amplitude Condutividades Térmicas	Desvio Padrão Ponderado Valências	Cobre	Desvio Padrão Ponderado Raios Atômicos	Amplitude Raios Atômicos	Entropia 1ª Energia Ionização	Média Ponderada Densidades	Desvio Padrão Condutividades Térmicas	Entropia Ponderada Calores de Fusão
LASSO	0,030	0,007	-0,042	0,021	-0,002	0,026	0,015	0,065	-0,026	0,033	0,018
Ridge	0,023	0,002	-0,026	0,011	-0,002	-0,004	0,009	-0,011	-0,025	0,020	0,012
Gama	0,073	0,051	0,033	0,018	0,024	0,031	0,035	0,014	0,010	0,025	0,010
Elastic Net	0,030	0,006	-0,035	0,015	-0,002	0,028	0,016	0,069	-0,028	0,025	0,016
Relaxed Lasso	0,028	0,007	-0,041	0,013	-0,002	0,026	0,017	0,076	-0,024	0,030	0,021
KNN	-0,013	-0,011	-0,008	-0,008	-0,009	-0,010	-0,009	-0,013	-0,006	-0,009	-0,039
SVM	0,077	0,022	0,011	0,037	-0,022	0,041	0,007	0,018	0,005	0,023	0,007
Decision Tree	0,179	0,266	0	0	0	0	0	0	0,103	0	0
Random Forest	0,028	0,004	0,017	0,038	-0,028	0,006	0,004	0,002	0,010	-0,001	0,003
Extra Trees	0,147	0,076	0,143	0,015	0,026	0,006	0,059	0,001	0,005	0,039	-0,001
XGBoost	0,041	0,051	0,026	0,036	0,024	0,036	0,014	0,011	0,016	0,033	0,005
LightGBM	0,110	0,011	0,025	0,040	0,051	0,027	0,043	-0,011	0,020	0,015	0,007
CatBoost	0,044	0,024	0,008	0,024	0,046	0,030	0,027	0,003	0,022	0,004	0,057
MLP	0,104	0,009	0,014	0,019	-0,016	0,003	-0,001	-0,000	0,003	0,008	-0,031
ResNet	0,023	0,004	0,017	0,020	-0,030	-0,018	0,036	0,013	0,015	0,031	0,009
TabNet	0,013	0,014	0,006	0,045	-0,034	0,035	0,013	0,006	0,002	0,014	0,051
FastFormer	-0,003	-0,016	-0,013	0,019	-0,044	-0,004	0,007	-0,001	0,011	-0,009	-0,023
FT-Transformer	0,055	0,029	0,054	0,029	0,026	0,015	0,009	0,011	0,000	0,025	0,006
TabTransformer	-0,002	0,000	0,001	-0,001	-0,001	-0,015	-0,029	-0,023	-0,016	-0,001	-0,029
Média	0,052	0,029	0,010	0,021	0,000	0,014	0,014	0,012	0,005	0,016	0,005
Desvio Padrão	0,050	0,060	0,040	0,014	0,026	0,018	0,019	0,027	0,028	0,015	0,024

Tabela 21 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Supercondutores

Modelo	Amplitude Condutividades Térmicas	Cobre	Bário	Desvio Padrão Condutividades Térmicas	Desvio Padrão Ponderado Condutividades Térmicas	Entropia Ponderada Massas Atômicas	Desvio Padrão Raios Atômicos	Média Ponderada Massas Atômicas	Desvio Padrão Afinidades Eletrônicas	Média Ponderada Condutividades Térmicas
LASSO	-0,025	-0,002	0,012	0,025	0,004	0,008	-0,024	-0,036	0,039	0,034
Ridge	-0,019	-0,001	0,009	0,021	0,001	0,002	-0,008	-0,044	0,028	0,026
Gama	0,018	0,015	0,029	0,017	0,025	0,014	0,013	-0,010	0,003	0,012
Elastic Net	-0,025	-0,002	0,012	0,025	0,004	0,008	-0,024	-0,036	0,039	0,034
Relaxed Lasso	-0,025	-0,002	0,012	0,025	0,004	0,008	-0,024	-0,036	0,039	0,034
Decision Tree	0,556	0,024	0,017	0,008	0,000	0,000	0,001	0,001	0,000	0,000
Random Forest	0,022	0,655	0,034	0,001	0,003	0,001	0,001	0,003	0,002	0,006
Extra Trees	0,221	0,031	0,049	0,067	0,148	0,028	0,003	0,003	0,002	0,010
XGBoost	0,036	0,030	0,051	0,012	0,049	0,015	0,001	0,002	0,010	0,008
LightGBM	0,005	0,004	0,004	0,008	0,005	0,005	0,006	0,011	0,016	0,016
CatBoost	0,004	0,042	0,021	0,010	0,021	0,018	0,009	0,011	0,020	0,012
TabNet	0	0,000	0,074	0,047	0,000	0,150	0,094	0,008	0,000	0
Média Absolutos	0,080	0,067	0,027	0,022	0,022	0,021	0,017	0,017	0,016	0,016
Desvio Padrão dos Absolutos	0,155	0,178	0,020	0,018	0,040	0,040	0,025	0,015	0,015	0,012

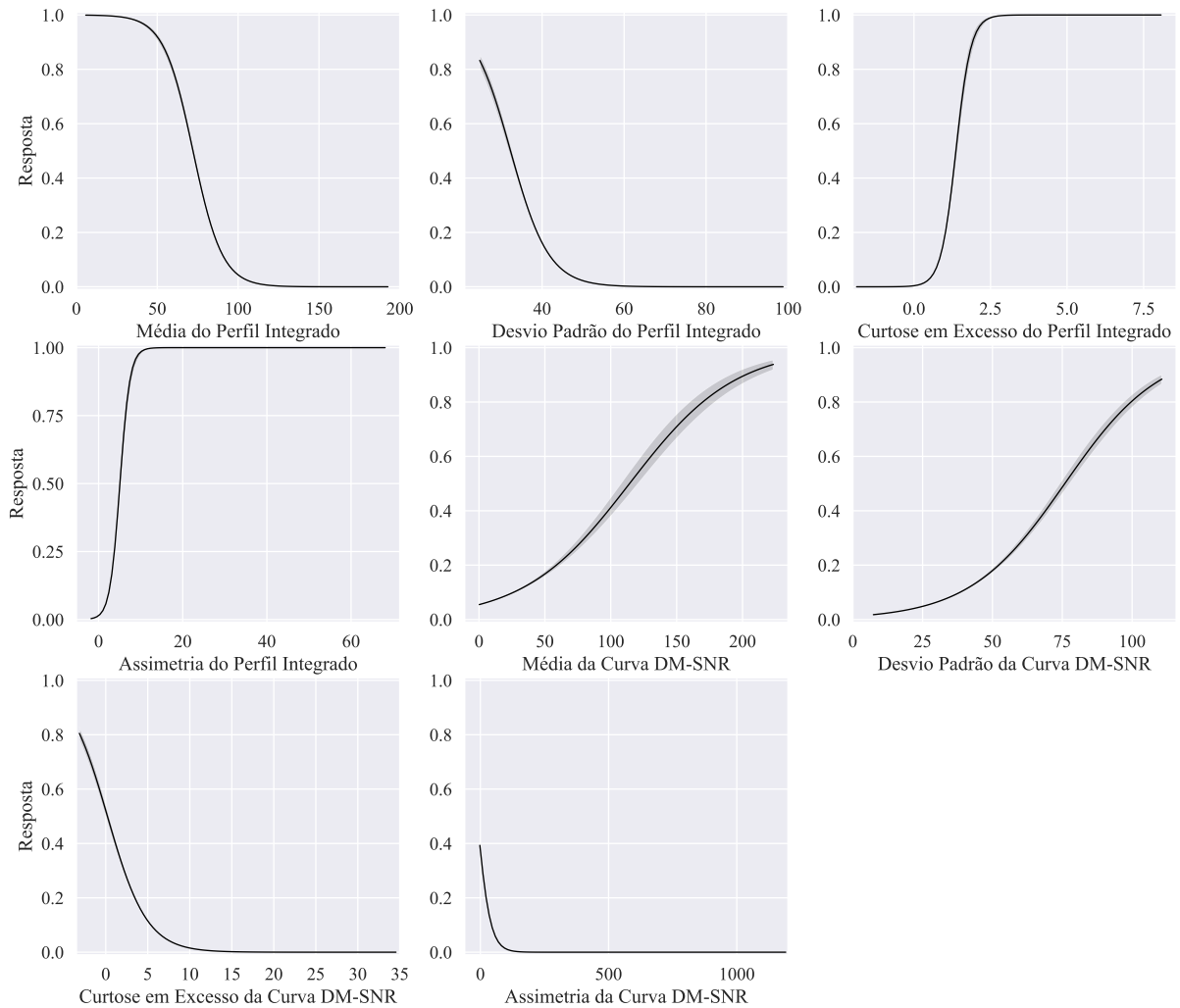
4.5 IDENTIFICAÇÃO DE PULSARES

Figura 35 – Histogramas das Variáveis para a Base de Dados das Pulsares



Fonte: Figura do Autor

Figura 36 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados das Pulsares



Fonte: Figura do Autor

Tabela 22 – Métricas para a Base de Dados das Pulsares

Modelo	Acurácia Balançada	F0,5	F1	F2	AUROC	Recall	Precisão	AUPR	Log Loss	Brier Score	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,950	0,972	0,971	0,970	0,981	0,969	0,973	0,939	0,064	0,015	0,831	0,835	0,81	20
Ridge	0,950	0,972	0,970	0,969	0,981	0,969	0,973	0,939	0,064	0,015	0,829	0,833	0,71	17
Elastic Net	0,950	0,972	0,971	0,970	0,981	0,970	0,973	0,939	0,064	0,016	0,833	0,836	0,93	23
KNN	0,943	0,968	0,967	0,966	0,976	0,965	0,970	0,931	0,081	0,017	0,810	0,814	0,04	1
SVM	0,944	0,976	0,975	0,975	0,981	0,975	0,976	0,936	0,065	0,016	0,854	0,855	21,89	532
Decision Tree	0,932	0,961	0,958	0,957	0,976	0,956	0,964	0,922	0,077	0,019	0,766	0,773	0,07	2
Random Forest	0,952	0,967	0,964	0,962	0,985	0,962	0,969	0,946	0,062	0,015	0,798	0,807	28,79	700
Extra Trees	0,955	0,976	0,975	0,975	0,984	0,975	0,977	0,946	0,060	0,015	0,857	0,859	3,80	92
XGBoost	0,956	0,975	0,973	0,973	0,980	0,972	0,975	0,942	0,066	0,016	0,846	0,849	0,69	17
LightGBM	0,955	0,974	0,972	0,972	0,986	0,971	0,975	0,948	0,059	0,015	0,841	0,844	1,14	28
CatBoost	0,952	0,971	0,970	0,969	0,979	0,968	0,973	0,941	0,067	0,016	0,826	0,831	5,08	123
MLP	0,954	0,980	0,979	0,979	0,984	0,979	0,980	0,943	0,062	0,015	0,879	0,879	25,72	625
ResNet	0,952	0,971	0,970	0,969	0,984	0,968	0,973	0,939	0,124	0,028	0,826	0,831	35,20	855
TabNet	0,946	0,966	0,964	0,962	0,985	0,962	0,968	0,935	0,142	0,033	0,796	0,803	36,27	881
SAINT	0,945	0,970	0,969	0,968	0,985	0,967	0,971	0,941	0,075	0,017	0,819	0,823	202,83	4927
FastFormer	0,926	0,954	0,949	0,945	0,970	0,945	0,958	0,905	0,153	0,042	0,719	0,732	28,37	689
FT-Transformer	0,955	0,974	0,972	0,972	0,984	0,971	0,975	0,945	0,131	0,029	0,841	0,844	31,86	774
TabTransformer	0,950	0,976	0,975	0,975	0,987	0,975	0,976	0,943	0,114	0,028	0,855	0,857	17,24	419
Trivial	0,5	0,841	0,865	0,890	0,5	0,908	0,825	0,546	0,306	0,083	0	0	-	-

Tabela 23 – SHAP Médio para a Base de Dados das Pulsares

Modelo	Curtose em Excesso do Perfil Integrado	Assimetria do Perfil Integrado	Desvio Padrão da Curva DM-SNR	Média da Curva DM-SNR	Média do Perfil Integrado	Desvio Padrão do Perfil Integrado	Assimetria da Curva DM-SNR	Curtose em Excesso da Curva DM-SNR
LASSO	0,714	0,131	0,096	-0,033	-0,008	-0,001	0,018	0
Ridge	0,710	0,128	0,098	-0,033	-0,009	-0,001	0,020	-0,002
Elastic Net	0,718	0,121	0,098	-0,034	-0,010	-0,000	0,019	0
KNN	0,355	0,217	0,114	0,018	0,119	-0,030	0,057	0,088
SVM	0,736	0,131	0,081	-0,024	0,015	-0,002	0,007	-0,003
Decision Tree	0,403	0,597	0	0	0	0	0	0
Random Forest	0,661	-0,032	-0,052	-0,086	-0,069	-0,046	0,026	0,028
Extra Trees	0,406	-0,153	-0,111	-0,141	0,042	-0,073	0,035	0,040
XGBoost	0,368	-0,076	0,050	-0,041	-0,113	0,078	0,208	0,065
LightGBM	0,479	-0,166	-0,083	-0,051	-0,102	-0,049	0,033	0,036
CatBoost	0,189	-0,200	0,144	0,156	0,151	0,146	0,013	0,002
MLP	0,474	0,070	0,029	-0,133	-0,015	-0,182	-0,015	-0,082
ResNet	-0,587	0,168	0,042	-0,017	0,038	-0,099	0,036	0,013
TabNet	0,661	0,000	0,002	-0,104	-0,100	-0,043	-0,069	-0,020
SAINT	-0,472	-0,031	-0,145	-0,028	-0,148	-0,133	0,037	-0,006
FastFormer	-0,185	0,067	0,030	-0,197	0,019	0,069	0,221	0,212
FT-Transformer	-0,390	-0,039	-0,176	-0,047	-0,181	-0,094	0,038	-0,036
TabTransformer	-0,148	0,168	-0,213	-0,115	-0,096	-0,189	0,022	0,049
Média	0,283	0,061	0,000	-0,051	-0,026	-0,036	0,039	0,021
Desvio Padrão	0,432	0,178	0,104	0,074	0,085	0,084	0,067	0,059

Tabela 24 – SHAP Absolutos para a Base de Dados das Pulsares

Modelo	Curtose em Excesso do Perfil Integrado	Desvio Padrão da Curva DM-SNR	Assimetria do Perfil Integrado	Média do Perfil Integrado	Desvio Padrão do Perfil Integrado	Média da Curva DM-SNR	Curtose em Excesso da Curva DM-SNR	Assimetria da Curva DM-SNR
LASSO	0,615	0,109	0,107	0,060	0,023	0,063	0	0,023
Ridge	0,611	0,111	0,106	0,059	0,023	0,063	0,002	0,025
Elastic Net	0,619	0,113	0,104	0,054	0,022	0,064	0	0,024
KNN	0,319	0,145	0,162	0,168	0,061	0,023	0,069	0,053
SVM	0,614	0,100	0,089	0,096	0,023	0,061	0,006	0,012
Decision Tree	0,617	0	0,383	0	0	0	0	0
Random Forest	0,647	0,190	0,056	0,028	0,022	0,032	0,016	0,010
Extra Trees	0,431	0,189	0,175	0,071	0,038	0,054	0,023	0,019
XGBoost	0,407	0,168	0,070	0,060	0,062	0,034	0,106	0,093
LightGBM	0,468	0,214	0,125	0,052	0,055	0,029	0,041	0,017
CatBoost	0,374	0,211	0,107	0,082	0,092	0,071	0,044	0,018
MLP	0,366	0,202	0,147	0,072	0,084	0,053	0,032	0,045
ResNet	0,352	0,161	0,104	0,176	0,060	0,079	0,027	0,040
TabNet	0,661	0,002	0,000	0,100	0,043	0,104	0,020	0,069
SAINT	0,310	0,214	0,128	0,130	0,093	0,027	0,034	0,065
FastFormer	0,060	0,102	0,067	0,347	0,071	0,057	0,163	0,132
FT-Transformer	0,231	0,196	0,252	0,148	0,097	0,026	0,029	0,022
TabTransformer	0,197	0,197	0,266	0,153	0,052	0,029	0,082	0,022
Média	0,439	0,146	0,136	0,103	0,051	0,048	0,039	0,038
Desvio Padrão	0,174	0,065	0,086	0,076	0,028	0,024	0,042	0,033

Tabela 25 – Feature Permutation para a Base de Dados das Pulsares

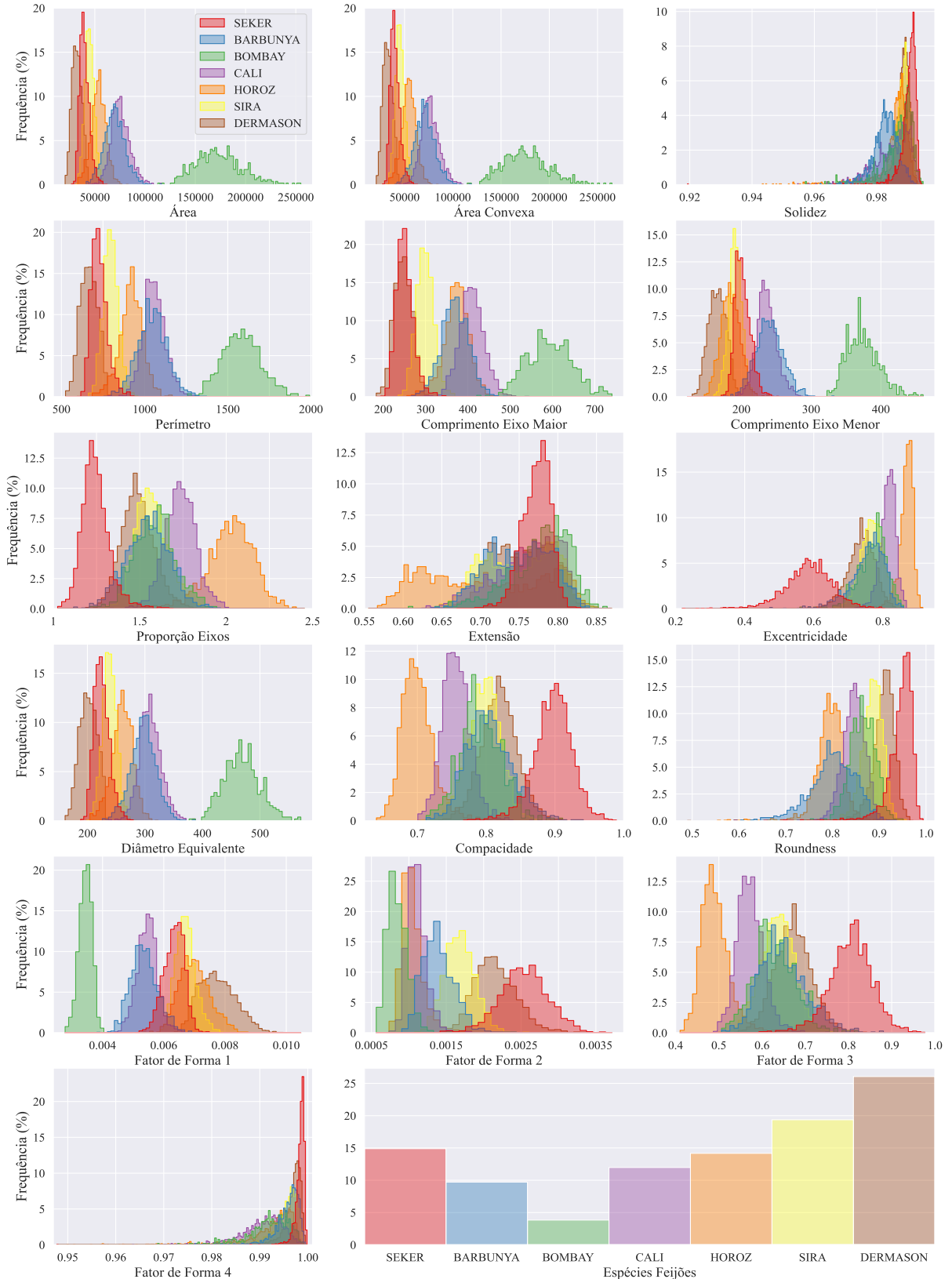
Modelo	Curtose em Excesso do Perfil Integrado	Assimetria do Perfil Integrado	Desvio Padrão da Curva DM-SNR	Média do Perfil Integrado	Desvio Padrão do Perfil Integrado	Média da Curva DM-SNR	Curtose em Excesso da Curva DM-SNR	Assimetria da Curva DM-SNR
LASSO	-0,881	-0,009	-0,077	0,019	0,005	0,004	0	-0,005
Ridge	-0,848	-0,017	-0,093	0,022	0,008	0,004	0,000	-0,008
Elastic Net	-0,834	-0,015	-0,108	0,023	0,006	0,005	0	-0,008
KNN	-0,566	-0,231	-0,082	-0,107	-0,007	-0,002	-0,002	-0,003
SVM	-0,783	-0,009	-0,156	0,032	0,006	0,008	0,001	-0,004
Decision Tree	-0,533	-0,467	0	0	0	0	0	0
Random Forest	-0,866	-0,038	-0,057	-0,008	0,001	-0,003	-0,004	-0,023
Extra Trees	-0,556	-0,173	-0,095	-0,124	-0,013	-0,019	-0,010	-0,010
XGBoost	-0,883	-0,008	-0,074	-0,007	-0,008	0,001	-0,001	-0,018
LightGBM	-0,757	-0,039	-0,146	-0,011	-0,013	-0,012	-0,007	-0,016
CatBoost	-0,270	-0,037	-0,425	-0,207	0,032	0,005	0,003	-0,022
MLP	-0,920	-0,028	-0,029	-0,014	-0,001	-0,002	0,002	0,005
ResNet	-0,253	0,337	-0,007	0,126	-0,087	-0,123	0,024	0,043
TabNet	0,475	0,041	0,064	0,274	-0,006	0,019	0,114	0,007
SAINT	-0,180	0,181	0,184	-0,132	-0,241	0,001	0,068	0,015
FastFormer	-0,130	-0,151	-0,088	-0,182	-0,107	-0,125	0,090	0,128
FT-Transformer	-0,234	0,023	-0,019	-0,210	-0,498	-0,002	-0,013	0,001
TabTransformer	-0,277	-0,012	-0,080	-0,173	-0,111	-0,181	0,117	0,050
Média	-0,517	-0,036	-0,072	-0,038	-0,057	-0,023	0,021	0,007
Desvio Padrão	0,366	0,159	0,115	0,121	0,125	0,055	0,042	0,035

Tabela 26 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados das Pulsares

Modelo	Curtose em Excesso do Perfil Integrado	Assimetria do Perfil Integrado	Média do Perfil Integrado	Desvio Padrão da Curva DM-SNR	Desvio Padrão do Perfil Integrado	Média da Curva DM-SNR	Assimetria da Curva DM-SNR	Curtose em Excesso da Curva DM-SNR
LASSO	0,497	-0,267	0,054	0,075	-0,019	-0,066	-0,024	0
Ridge	0,494	-0,265	0,053	0,076	-0,019	-0,066	-0,025	0,002
Elastic Net	0,498	-0,263	0,050	0,078	-0,018	-0,068	-0,025	0
Decision Tree	0,136	0,864	0	0	0	0	0	0
Random Forest	0,845	0,079	0,010	0,036	0,009	0,009	0,005	0,006
Extra Trees	0,394	0,179	0,210	0,101	0,022	0,033	0,023	0,038
XGBoost	0,403	0,200	0,157	0,051	0,048	0,052	0,042	0,049
LightGBM	0,172	0,133	0,148	0,130	0,140	0,100	0,088	0,089
CatBoost	0,211	0,116	0,150	0,140	0,184	0,074	0,062	0,063
TabNet	0,283	0	0,400	0,098	0,148	0,028	0,015	0,028
Média Absolutos	0,393	0,237	0,123	0,079	0,061	0,050	0,031	0,028
Desvio Padrão dos Absolutos	0,199	0,225	0,114	0,040	0,065	0,030	0,025	0,030

4.6 IDENTIFICAÇÃO DE ESPÉCIES DE FEIJÕES

Figura 37 – Histogramas das Variáveis para a Base de Dados dos Feijões



Fonte: Figura do Autor

A Variável Diâmetro Equivalente descreve o diâmetro de um círculo com a mesma área de um dado grão. A variável Compacidade descreve o quão próximo o formato do grão está de um círculo, sendo calculado pela razão entre o Diâmetro Equivalente e o Comprimento do Eixo Maior. Já as variáveis *Roundness*, Fator de Forma indo do número 1 até 4; são funções das dimensões, área e perímetro do grão e estão descritas no artigo original de Koklu e Ozkan (2020).

A variável Comprimento Eixo Maior trata da dimensão da maior linha reta contínua sobre o grão. A variável Comprimento Eixo Menor é a dimensão da maior linha reta perpendicular ao eixo Maior, tendo a Proporção Eixos como a razão do Eixo Maior pelo Menor. Sendo que a Área Convexa representa a área do fecho convexo que contém a do grão, com a variável Solidez sendo a razão entre a área do grão e a desse fecho convexo.

Tabela 27 – Métricas para a Base de Dados dos Feijões

Modelo	Acurácia	F0,5	F1	F2	AUROC	Recall	Precisão	Log Loss	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,918	0,918	0,918	0,918	0,993	0,918	0,918	0,217	0,901	0,901	23,97	373
Ridge	0,916	0,916	0,916	0,916	0,993	0,916	0,916	0,214	0,899	0,899	7,70	120
Elastic Net	0,918	0,918	0,918	0,918	0,993	0,918	0,918	0,217	0,901	0,901	25,75	401
KNN	0,912	0,913	0,912	0,912	0,992	0,912	0,913	0,325	0,894	0,894	0,06	1
SVM	0,926	0,925	0,925	0,925	0,994	0,926	0,926	0,200	0,910	0,910	5,87	91
Decision Tree	0,889	0,889	0,888	0,889	0,989	0,889	0,889	0,395	0,866	0,866	0,16	2
Random Forest	0,915	0,915	0,915	0,915	0,993	0,915	0,915	0,228	0,898	0,898	68,24	1062
Extra Trees	0,914	0,914	0,914	0,914	0,993	0,914	0,914	0,217	0,896	0,896	5,82	91
XGBoost	0,921	0,921	0,921	0,921	0,994	0,921	0,921	0,205	0,905	0,905	1,58	25
LightGBM	0,925	0,925	0,925	0,925	0,994	0,925	0,925	0,190	0,909	0,910	3,41	53
CatBoost	0,927	0,927	0,927	0,927	0,994	0,927	0,927	0,204	0,912	0,912	78,84	1227
MLP	0,906	0,906	0,906	0,906	0,993	0,906	0,907	0,227	0,887	0,887	18,81	293
ResNet	0,933	0,933	0,933	0,933	0,995	0,933	0,934	0,196	0,919	0,919	43,07	670
TabNet	0,915	0,916	0,915	0,915	0,993	0,915	0,918	0,239	0,897	0,898	49,84	776
SAINT	0,924	0,925	0,924	0,924	0,994	0,924	0,925	0,209	0,908	0,909	13,97	217
FastFormer	0,932	0,932	0,931	0,932	0,994	0,932	0,932	0,196	0,918	0,918	30,40	473
FT-Transformer	0,931	0,931	0,931	0,931	0,995	0,931	0,931	0,187	0,917	0,917	22,71	353
TabTransformer	0,923	0,924	0,923	0,923	0,994	0,923	0,925	0,203	0,907	0,908	32,08	499
Trivial	0,261	0,080	0,108	0,166	0,5	0,261	0,068	1,834	0,887	0,887	-	-

É retratado a importância para somente uma das classes: feijão Barbunya, cultivar turco similar ao feijão-rajado.

Tabela 28 – SHAP Médio para a Base de Dados dos Feijões

Modelo	Roundness	Fator de Forma 2	Fator de Forma 1	Perímetro	Fator de Forma 3	Comprimento Eixo Menor	Comprimento Eixo Maior	Solidez	Área Convexa	Proporção Eixos
LASSO	0,052	0,306	-0,086	-0,031	-0,022	0,173	-0,108	0,074	-0,038	-0,006
Ridge	0,001	0,398	-0,104	0,025	0,002	-0,085	-0,046	0,048	0,025	0,095
Elastic Net	0,056	0,287	-0,086	-0,025	-0,026	0,178	-0,111	0,078	-0,037	-0,011
KNN	0,314	-0,025	0,115	0,081	-0,038	0,074	-0,009	0,086	0,012	-0,039
SVM	0,289	-0,038	0,122	0,089	-0,059	0,058	-0,014	0,049	0,005	-0,059
Decision Tree	0,295	-0,003	0,000	-0,010	-0,224	-0,024	-0,034	-0,000	0,330	-0,002
Random Forest	0,314	-0,085	0,095	0,238	-0,050	0,014	-0,097	0,001	-0,004	-0,005
Extra Trees	0,275	-0,141	0,074	0,171	-0,077	0,029	-0,041	0,003	0,005	-0,035
XGBoost	0,291	-0,026	0,128	0,146	-0,040	0,058	-0,025	0,029	0,009	-0,022
LightGBM	0,295	-0,059	0,182	0,054	-0,057	0,073	0,007	-0,010	0,001	-0,049
CatBoost	0,276	-0,086	0,171	0,117	-0,052	0,029	-0,026	-0,035	0,023	-0,031
MLP	0,194	-0,117	0,111	0,047	-0,077	0,076	-0,051	0,061	0,041	-0,026
ResNet	-0,128	0,054	-0,161	-0,032	-0,123	-0,086	0,042	0,029	0,090	0,009
SAINT	0,023	-0,044	-0,014	0,025	-0,077	-0,090	-0,105	-0,160	-0,041	-0,088
FastFormer	0,139	-0,093	-0,058	-0,074	-0,048	-0,003	-0,093	-0,096	-0,025	-0,056
FT-Transformer	0,061	-0,204	0,004	-0,038	-0,136	-0,035	0,005	-0,075	-0,078	-0,149
TabTransformer	0,027	-0,081	0,014	-0,049	-0,102	-0,072	-0,070	-0,022	-0,092	-0,101
Média	0,163	0,002	0,030	0,043	-0,071	0,022	-0,046	0,004	0,013	-0,034
Desvio Padrão	0,138	0,163	0,101	0,085	0,051	0,080	0,044	0,065	0,090	0,051

Tabela 29 – SHAP Absolutos para a Base de Dados dos Feijões

Modelo	Roundness	Fator de Forma 4	Fator de Forma 1	Fator de Forma 2	Perímetro	Solidez	Comprimento Eixo Menor	Fator de Forma 3	Área Convexa	Proporção Eixos
LASSO	0,047	0,093	0,038	0,242	0,033	0,076	0,111	0,014	0,025	0,109
Ridge	0,019	0,046	0,038	0,236	0,051	0,030	0,050	0,035	0,091	0,193
Elastic Net	0,050	0,097	0,038	0,228	0,034	0,081	0,115	0,015	0,026	0,103
KNN	0,219	0,126	0,089	0,023	0,063	0,107	0,061	0,046	0,023	0,049
SVM	0,193	0,146	0,085	0,025	0,070	0,092	0,053	0,048	0,028	0,050
Decision Tree	0,267	0,047	0,002	0,004	0,010	0,001	0,023	0,203	0,309	0,003
Random Forest	0,201	0,068	0,073	0,049	0,206	0,034	0,045	0,077	0,027	0,019
Extra Trees	0,201	0,037	0,063	0,087	0,181	0,023	0,041	0,069	0,042	0,040
XGBoost	0,222	0,120	0,097	0,026	0,112	0,070	0,065	0,035	0,017	0,024
LightGBM	0,198	0,173	0,110	0,032	0,050	0,088	0,067	0,041	0,030	0,030
CatBoost	0,198	0,149	0,118	0,049	0,088	0,062	0,048	0,040	0,031	0,029
MLP	0,159	0,105	0,091	0,070	0,066	0,078	0,080	0,066	0,068	0,018
ResNet	0,093	0,025	0,126	0,018	0,048	0,012	0,060	0,170	0,109	0,017
SAINT	0,146	0,121	0,072	0,017	0,072	0,113	0,066	0,065	0,037	0,057
FastFormer	0,128	0,069	0,158	0,053	0,064	0,092	0,078	0,037	0,028	0,038
FT-Transformer	0,133	0,139	0,139	0,062	0,034	0,110	0,071	0,041	0,023	0,087
TabTransformer	0,170	0,118	0,094	0,030	0,064	0,069	0,060	0,064	0,044	0,064
Média	0,155	0,099	0,084	0,074	0,073	0,067	0,064	0,063	0,056	0,055
Desvio Padrão	0,067	0,042	0,039	0,078	0,049	0,034	0,022	0,049	0,068	0,045

Tabela 30 – Feature Permutation para a Base de Dados dos Feijões

Modelo	Roundness	Fator de Forma 3	Fator de Forma 1	Área Convexa	Perímetro	Solidez	Fator de Forma 4	Proporção Eixos	Excetricidade	Comprimento Eixo Maior
LASSO	0,116	-0,030	0,127	0,082	0,121	0,077	-0,058	-0,041	0,002	-0,026
Ridge	0,092	0,077	0,079	0,064	0,114	0,052	-0,023	-0,026	-0,030	0,103
Elastic Net	0,119	-0,027	0,135	0,087	0,127	0,090	-0,036	-0,024	0,007	-0,015
KNN	0,104	-0,099	0,067	0,022	0,065	0,093	-0,102	-0,091	-0,075	0,029
SVM	0,205	-0,049	0,101	0,040	0,085	0,128	-0,016	-0,046	-0,042	0,044
Decision Tree	-0,162	-0,434	0,000	-0,308	0,000	0,000	0,000	0,000	0,000	0,077
Random Forest	0,093	-0,162	0,026	0,014	0,089	0,019	-0,036	-0,041	-0,033	-0,187
Extra Trees	0,182	-0,098	0,059	0,048	0,107	0,030	-0,053	-0,048	-0,040	0,020
XGBoost	0,346	0,000	0,080	0,022	0,088	0,227	-0,007	0,032	0,003	0,048
LightGBM	0,285	-0,025	0,185	0,045	0,031	0,190	0,000	0,001	0,001	0,023
CatBoost	0,129	-0,054	0,086	0,035	0,053	0,119	-0,076	-0,044	-0,059	0,027
MLP	0,148	-0,133	0,078	0,042	0,085	0,064	-0,096	0,006	-0,082	-0,050
ResNet	0,051	0,246	0,060	0,167	-0,010	-0,001	-0,015	-0,007	0,261	-0,008
SAINT	0,029	-0,153	-0,016	-0,050	-0,040	-0,001	-0,256	-0,140	-0,088	-0,071
FastFormer	-0,029	-0,089	-0,086	-0,039	-0,064	0,000	-0,131	-0,079	-0,087	-0,098
FT-Transformer	-0,218	-0,028	0,010	-0,056	0,039	-0,044	-0,119	-0,277	-0,022	-0,019
TabTransformer	0,106	-0,103	-0,016	-0,076	-0,050	0,004	-0,113	-0,098	-0,085	-0,069
Média	0,094	-0,068	0,057	0,008	0,049	0,062	-0,067	-0,054	-0,022	-0,010
Desvio Padrão	0,135	0,130	0,064	0,098	0,060	0,071	0,064	0,070	0,078	0,068

Tabela 31 – Importância das Variáveis Internamente nos Modelos Lineares para a Base de Dados dos Feijões

Modelo	Fator de Forma 2	Comprimento Eixo Menor	Proporção Eixos	Fator de Forma 4	Roundness	Área Convexa	Comprimento Eixo Maior	Excetricidade	Solidez	Diâmetro Equivalente
LASSO	-0,256	0,160	-0,106	0,092	-0,078	0,021	-0,078	-0,040	-0,065	0,037
Ridge	-0,198	-0,088	-0,147	0,037	-0,030	0,145	0,028	-0,099	-0,022	-0,050
Elastic Net	-0,243	0,156	-0,103	0,096	-0,079	0,021	-0,071	-0,038	-0,068	0,039
Média	-0,232	0,076	-0,118	0,075	-0,062	0,062	-0,040	-0,059	-0,052	0,008
Desvio Padrão	0,025	0,116	0,020	0,027	0,023	0,058	0,049	0,028	0,021	0,041

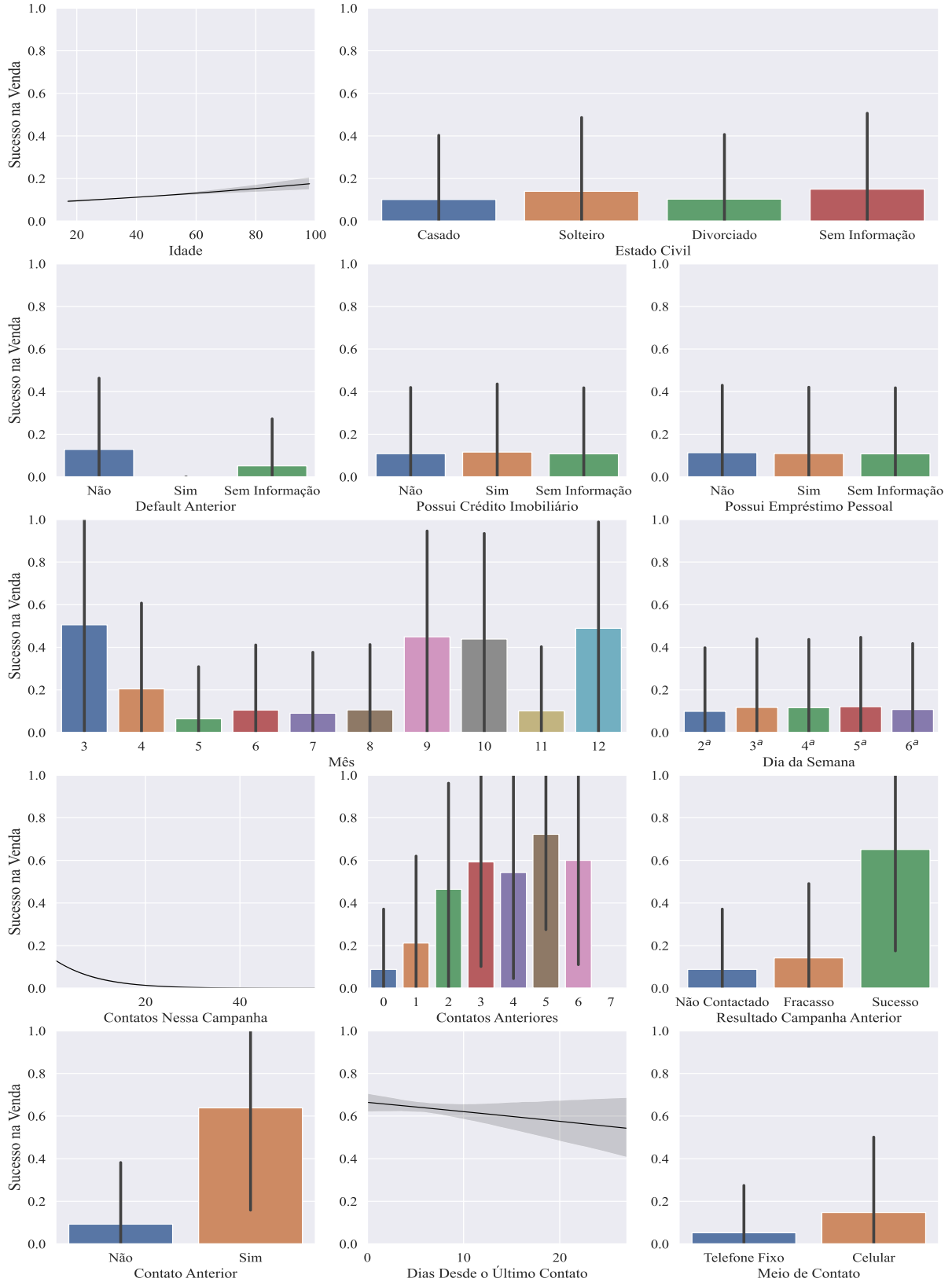
Já a importância interna para os métodos baseados em Árvores de Decisão e TabNet são mais gerais, não dependendo da classe escolhida.

Tabela 32 – Importância das Variáveis Internamente para a Base de Dados dos Feijões

Modelo	Fator de Forma 1	Fator de Forma 3	Comprimento Eixo Menor	Comprimento Eixo Maior	Compacidade	Roundness	Área Convexa	Perímetro	Fator de Forma 4	Área
Decision Tree	0,195	0,221	0,096	0,206	0,039	0,059	0,153	0,014	0,010	0,000
Random Forest	0,140	0,081	0,117	0,159	0,078	0,040	0,050	0,134	0,021	0,030
Extra Trees	0,097	0,088	0,077	0,070	0,090	0,056	0,060	0,087	0,018	0,061
XGBoost	0,030	0,104	0,066	0,038	0,101	0,027	0,075	0,043	0,019	0,128
LightGBM	0,048	0,055	0,048	0,034	0,056	0,092	0,036	0,036	0,122	0,036
CatBoost	0,072	0,029	0,067	0,056	0,032	0,121	0,036	0,054	0,129	0,028
TabNet	0,091	0,075	0,144	0,028	0,180	0,081	0,039	0,059	0,057	0,043
Média	0,096	0,093	0,088	0,084	0,082	0,068	0,064	0,061	0,054	0,047
Desvio Padrão	0,052	0,057	0,031	0,065	0,046	0,030	0,038	0,036	0,047	0,037

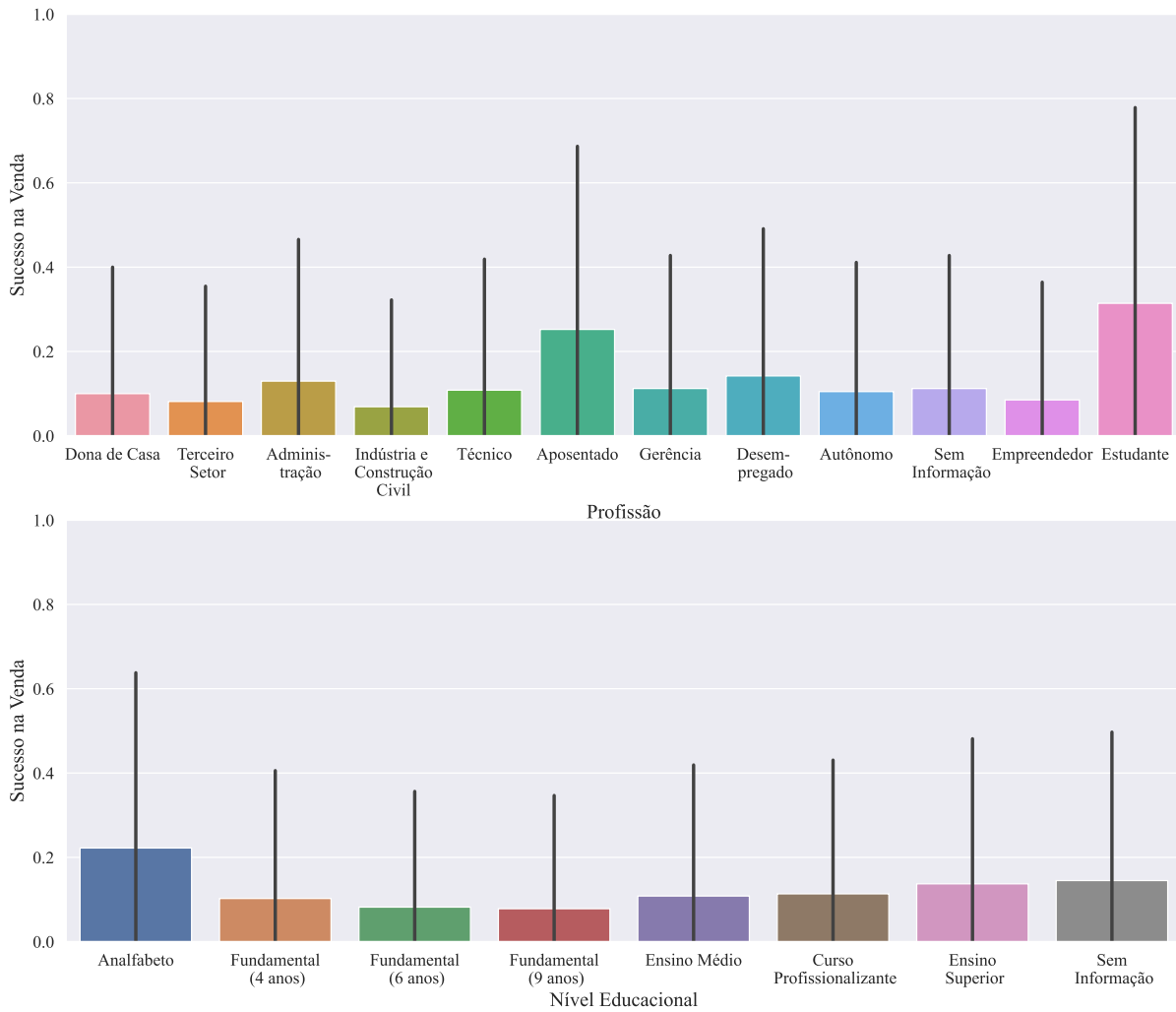
4.7 TELEMARKETING BANCO

Figura 38 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Marketing



Fonte: Figura do Autor

Figura 39 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Marketing



Fonte: Figura do Autor

Tabela 33 – Métricas para a Base de Dados do Marketing

Modelo	Acurácia Balançada	F0,5	F1	F2	AUROC	Recall	Precisão	AUPR	Log Loss	Brier Score	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,694	0,701	0,700	0,701	0,749	0,703	0,702	0,734	0,581	0,198	0,392	0,396	6,006	3392
Ridge	0,694	0,702	0,700	0,702	0,750	0,704	0,704	0,735	0,580	0,198	0,393	0,398	1,960	1106
Elastic Net	0,695	0,701	0,701	0,702	0,750	0,703	0,702	0,735	0,580	0,198	0,394	0,396	1,630	920
KNN	0,679	0,694	0,685	0,689	0,720	0,698	0,709	0,703	0,604	0,208	0,370	0,391	0,002	1
SVM	0,703	0,716	0,710	0,712	0,737	0,718	0,725	0,711	0,588	0,200	0,416	0,430	31,726	17914
Decision Tree	0,697	0,704	0,703	0,704	0,755	0,706	0,706	0,749	0,574	0,196	0,398	0,402	0,024	14
Random Forest	0,705	0,719	0,712	0,714	0,762	0,721	0,731	0,751	0,569	0,193	0,421	0,438	5,892	3327
Extra Trees	0,708	0,715	0,714	0,715	0,759	0,717	0,717	0,748	0,572	0,194	0,422	0,425	5,596	3160
XGBoost	0,712	0,726	0,719	0,721	0,766	0,728	0,737	0,753	0,567	0,191	0,435	0,451	0,595	336
LightGBM	0,695	0,707	0,702	0,704	0,750	0,709	0,715	0,737	0,634	0,205	0,399	0,411	0,846	478
CatBoost	0,715	0,723	0,721	0,722	0,769	0,725	0,726	0,753	0,567	0,191	0,436	0,441	9,501	5365
MLP	0,687	0,693	0,693	0,694	0,733	0,695	0,694	0,721	0,652	0,214	0,378	0,380	49,753	28093
ResNet	0,705	0,712	0,711	0,712	0,760	0,714	0,714	0,745	0,576	0,196	0,415	0,418	35,623	20114
TabNet	0,690	0,703	0,696	0,699	0,740	0,706	0,715	0,729	0,597	0,203	0,390	0,406	32,362	18273
SAINT	0,692	0,702	0,699	0,701	0,750	0,705	0,707	0,736	0,586	0,198	0,392	0,400	17,555	9913
FastFormer	0,692	0,703	0,699	0,701	0,752	0,706	0,711	0,732	0,613	0,198	0,393	0,404	45,181	25511
FT-Transformer	0,698	0,709	0,705	0,707	0,749	0,712	0,716	0,736	0,582	0,198	0,405	0,416	60,001	33879
TabTransformer	0,699	0,709	0,706	0,708	0,755	0,711	0,714	0,738	0,588	0,198	0,406	0,414	32,526	18365
Trivial	0,5	0,332	0,390	0,472	0,5	0,550	0,302	0,725	0,688	0,248	0	0	-	-

Tabela 34 – SHAP Médio para a Base de Dados do Marketing

Modelo	Sucesso Campanha Anterior	Contato Celular	Idade	Contato Últimos 7 Dias	Maio	Março	Sem Default Anterior	Outubro	Novembro	Julho
LASSO	-0,350	0,070	-0,002	-0,133	0,004	-0,084	0,016	-0,069	-0,011	-0,016
Ridge	-0,319	0,076	-0,002	-0,166	0,006	-0,085	0,020	-0,068	-0,012	-0,016
Elastic Net	-0,349	0,076	-0,001	-0,138	0,003	-0,097	0,017	-0,081	-0,008	-0,005
KNN	0,077	-0,046	0,034	0,062	0,001	0,060	-0,029	0,060	-0,014	0,025
SVM	-0,017	-0,013	0,078	0,001	0,031	0,016	0,037	0,014	0,012	0,016
Decision Tree	0,053	-0,115	0,015	-0,000	-0,200	0,007	-0,109	-0,000	-0,097	-0,114
Random Forest	0,117	-0,032	0,194	0,023	-0,110	0,033	-0,053	0,031	-0,033	-0,022
Extra Trees	-0,126	-0,175	0,087	-0,061	-0,120	-0,023	0,043	-0,001	-0,025	-0,023
XGBoost	0,092	-0,150	0,164	0,038	-0,054	0,028	-0,051	0,028	-0,028	-0,018
LightGBM	0,077	0,120	0,106	0,063	-0,048	0,032	0,151	0,029	0,014	-0,016
CatBoost	0,070	-0,164	0,124	0,027	-0,069	0,023	-0,046	0,024	-0,041	-0,034
MLP	-0,018	-0,043	0,200	-0,030	0,037	-0,003	-0,004	0,012	-0,062	0,028
ResNet	-0,003	0,064	-0,007	-0,010	0,034	-0,076	0,029	-0,017	0,013	0,045
TabNet	-0,036	-0,022	0,017	0,014	0,042	-0,065	0,003	-0,035	-0,027	-0,055
SAINT	-0,087	-0,025	0,009	-0,071	0,039	-0,060	0,068	-0,047	0,070	0,065
FastFormer	0,077	0,075	0,008	0,041	-0,013	-0,055	0,003	-0,058	0,092	0,021
FT-Transformer	-0,087	0,034	-0,046	0,022	0,021	-0,067	0,047	-0,048	0,002	0,001
TabTransformer	-0,068	0,062	-0,006	-0,077	0,039	-0,042	0,042	-0,049	0,060	0,091
Média	-0,050	-0,011	0,054	-0,022	-0,020	-0,025	0,010	-0,015	-0,005	-0,002
Desvio Padrão	0,147	0,089	0,073	0,069	0,066	0,049	0,055	0,041	0,045	0,045

Tabela 35 – SHAP Absolutos para a Base de Dados do Marketing

Modelo	Contato Celular	Sucesso Campanha Anterior	Maio	Idade	Sem Default Anterior	Novembro	Contato Últimos 7 Dias	Março	Julho	Junho
LASSO	0,112	0,079	0,045	0,005	0,043	0,030	0,029	0,027	0,023	0,021
Ridge	0,146	0,094	0,062	0,007	0,058	0,039	0,046	0,036	0,029	0,026
Elastic Net	0,163	0,114	0,032	0,004	0,063	0,027	0,044	0,044	0,010	0,048
KNN	0,103	0,045	0,025	0,023	0,025	0,014	0,037	0,035	0,041	0,031
SVM	0,089	0,089	0,067	0,028	0,038	0,035	0,027	0,028	0,029	0,067
Decision Tree	0,113	0,049	0,184	0,037	0,100	0,089	0,001	0,007	0,105	0,028
Random Forest	0,071	0,106	0,102	0,188	0,048	0,030	0,021	0,030	0,021	0,051
Extra Trees	0,120	0,167	0,112	0,022	0,079	0,021	0,044	0,038	0,031	0,069
XGBoost	0,134	0,082	0,049	0,157	0,046	0,025	0,034	0,025	0,017	0,054
LightGBM	0,120	0,055	0,035	0,118	0,109	0,013	0,045	0,024	0,016	0,011
CatBoost	0,142	0,061	0,060	0,124	0,040	0,036	0,024	0,020	0,033	0,039
MLP	0,082	0,048	0,059	0,058	0,031	0,040	0,024	0,020	0,029	0,030
ResNet	0,091	0,066	0,068	0,068	0,046	0,024	0,039	0,050	0,024	0,016
TabNet	0,022	0,036	0,042	0,017	0,003	0,027	0,014	0,065	0,055	0,056
SAINT	0,055	0,084	0,055	0,012	0,066	0,057	0,073	0,051	0,053	0,014
FastFormer	0,065	0,061	0,027	0,021	0,023	0,066	0,060	0,048	0,024	0,013
FT-Transformer	0,112	0,118	0,052	0,049	0,051	0,024	0,027	0,052	0,020	0,013
TabTransformer	0,069	0,065	0,067	0,019	0,039	0,053	0,057	0,031	0,062	0,015
Média	0,101	0,079	0,063	0,053	0,050	0,036	0,036	0,035	0,035	0,034
Desvio Padrão	0,035	0,032	0,037	0,055	0,026	0,019	0,017	0,014	0,022	0,019

Tabela 36 – Feature Permutation para a Base de Dados do Marketing

Modelo	Novembro	Sucesso Campanha Anterior	Contato Celular	Contato Últimos 7 Dias	Idade	Nível Educacional Curso Profissionalizante	Junho	Sem Default Anterior	Agosto	Casado
LASSO	-0,151	-0,075	0,069	-0,030	-0,003	-0,093	-0,021	0,021	0,025	0,079
Ridge	-0,203	-0,095	0,095	-0,057	-0,004	0,019	-0,027	0,035	0,018	0,031
Elastic Net	-0,152	-0,102	0,162	-0,049	0,000	0,006	-0,040	0,015	0,010	0,000
KNN	0,187	0,002	0,056	0,004	0,009	0,098	-0,007	0,038	0,013	0,078
SVM	-0,081	-0,076	0,015	-0,036	-0,003	0,161	-0,033	0,017	0,025	0,038
Decision Tree	-0,461	-0,090	-0,131	0,000	-0,053	0,000	-0,093	0,006	0,000	0,113
Random Forest	-0,049	-0,078	-0,094	-0,014	-0,165	0,004	-0,096	0,015	0,008	0,077
Extra Trees	-0,020	-0,053	-0,024	-0,015	-0,050	0,051	-0,125	0,059	0,026	0,078
XGBoost	-0,110	-0,095	0,017	-0,061	-0,140	0,050	-0,063	0,019	0,017	0,006
LightGBM	-0,019	-0,040	0,030	-0,029	-0,155	0,061	-0,040	0,017	0,013	0,028
CatBoost	-0,191	-0,083	-0,034	-0,031	-0,101	0,061	-0,016	0,015	0,056	0,001
MLP	0,065	-0,004	0,023	-0,000	-0,017	0,177	-0,008	0,035	0,028	0,051
ResNet	-0,034	-0,140	-0,186	0,016	0,051	-0,001	-0,016	0,054	-0,085	-0,001
TabNet	0,006	0,066	0,098	0,150	-0,017	0,017	0,026	0,046	0,012	-0,004
SAINT	-0,033	0,204	0,025	0,163	0,003	-0,001	-0,021	0,075	-0,055	-0,002
FastFormer	-0,032	0,197	0,003	0,178	0,054	0,002	0,001	0,038	-0,035	0,010
FT-Transformer	-0,061	0,077	-0,079	-0,088	-0,018	0,026	-0,011	0,042	-0,017	-0,002
TabTransformer	-0,100	0,072	-0,028	0,047	0,044	-0,004	-0,009	0,104	-0,195	0,019
Média	-0,080	-0,017	0,001	0,008	-0,031	0,035	-0,033	0,036	-0,008	0,033
Desvio Padrão	0,129	0,099	0,083	0,076	0,065	0,061	0,037	0,024	0,056	0,036

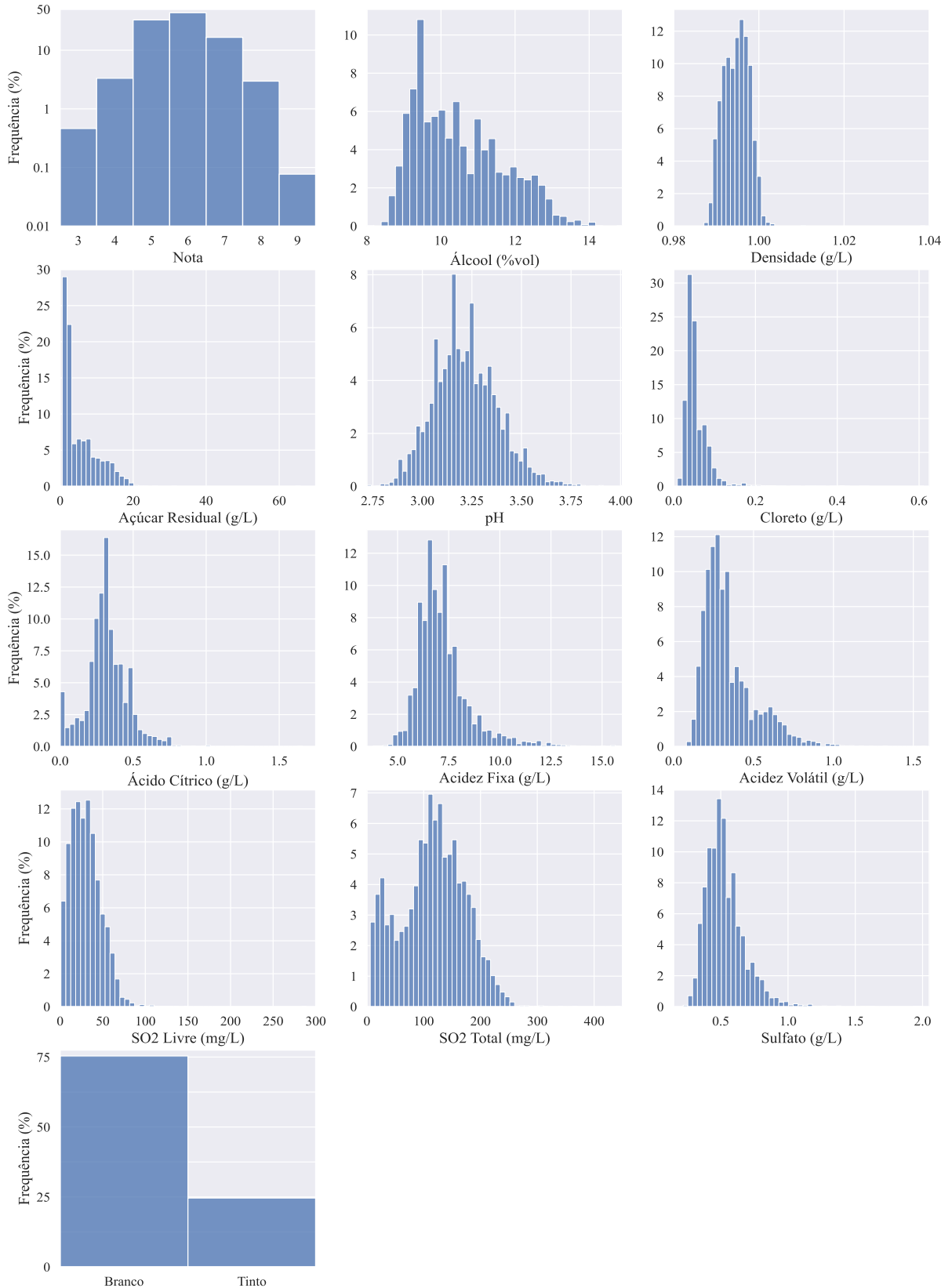
Tabela 37 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Marketing

Modelo	Sucesso Campanha Anterior	Contato Celular	Idade	Contato Últimos 7 Dias	Março	Outubro	Junho	Maior	Sem Default Anterior	Contatos Nessa Campanha
LASSO	0,095	0,078	0,004	0,038	0,054	0,052	0,020	-0,031	0,035	-0,022
Ridge	0,107	0,096	0,005	0,057	0,067	0,064	0,024	-0,040	0,045	-0,029
Elastic Net	0,125	0,102	0,003	0,052	0,079	0,078	0,042	-0,020	0,047	-0,029
Decision Tree	0,377	0,158	0,090	0	0,058	0	0,084	0,057	0,037	0,006
Random Forest	0,238	0,119	0,139	0,063	0,044	0,042	0,056	0,036	0,033	0,023
Extra Trees	0,245	0,128	0,031	0,073	0,050	0,052	0,059	0,038	0,041	0,006
XGBoost	0,246	0,067	0,010	0,171	0,066	0,048	0,040	0,014	0,025	0,007
LightGBM	0,008	0,024	0,179	0,009	0,006	0,007	0,015	0,022	0,019	0,115
CatBoost	0,107	0,093	0,166	0,039	0,026	0,030	0,032	0,041	0,028	0,060
TabNet	0,093	0,058	0,035	0,126	0,044	0,044	0,023	0,068	0,052	0,030
Média Absolutos	0,164	0,092	0,066	0,063	0,049	0,042	0,039	0,037	0,036	0,033
Desvio Padrão dos Absolutos	0,103	0,036	0,067	0,049	0,020	0,023	0,020	0,016	0,010	0,031

Interessante notar nas tabelas 34, 35, 36 e 37 que a profissão e o nível educacional do cliente foram considerados menos relevantes para a predição, além da grande relevância dada ao período do ano e o meio que o contato ocorreu.

4.8 NOTAS PARA VINHOS

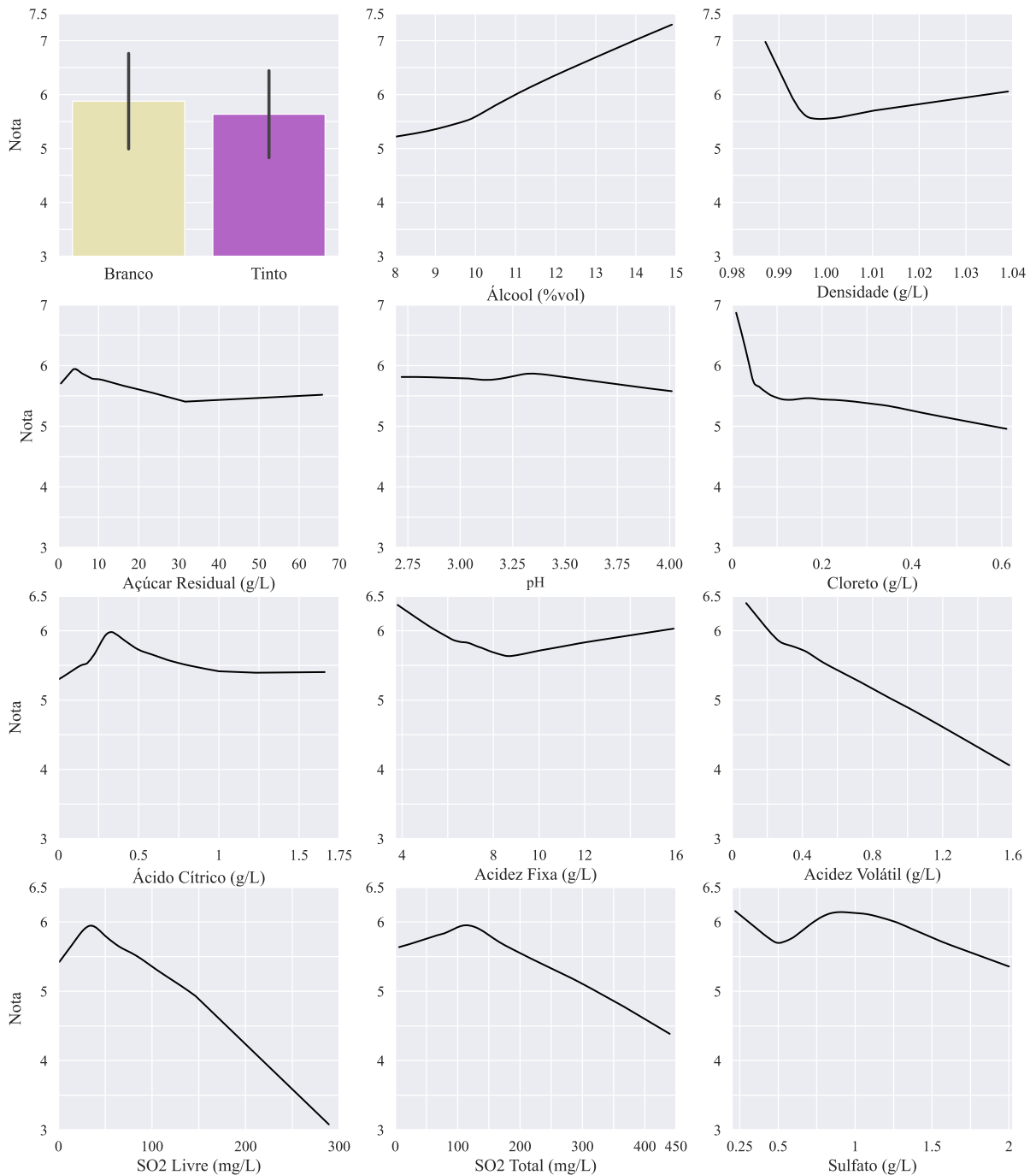
Figura 40 – Histogramas das Variáveis para a Base de Dados dos Vinhos



Fonte: Figura do Autor

A variável Cloreto está na unidade de grama de Cloreto de Sódio por litro de vinho. A variável Sulfato está na unidade de grama de Sulfato de Potássio por litro de vinho. A variável Acidez Volátil está na unidade de grama de Ácido Acético por litro de vinho. A variável Acidez Fixa está na unidade de grama de Ácido Tartárico por litro de vinho. A variável Álcool está na unidade de porcentagem da razão volumétrica entre o álcool etílico e o vinho.

Figura 41 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Vinhos



Fonte: Figura do Autor

Apesar das notas dadas para cada vinho serem números naturais foi escolhido realizar predições para valores contínuos, isso prejudica o valor nominal das métricas avaliadas, mas

permite uma maior granularidade na análise da importância dada para cada variável, além de discriminar quais vinhos são considerados melhores, mesmo que por uma pequena margem, por cada modelo.

Tabela 38 – Métricas para a Base de Dados dos Vinhos

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	0,521	0,567	3,357	0,314	0,314	0,039	6,4
Ridge	0,521	0,567	3,351	0,314	0,314	0,012	1,9
Elastic Net	0,521	0,567	3,350	0,314	0,314	0,023	3,8
Relaxed Lasso	0,521	0,567	3,356	0,314	0,314	0,030	5,0
KNN	0,424	0,483	3,590	0,441	0,442	0,018	3,0
SVM	0,453	0,511	3,340	0,404	0,407	2,480	410,1
Decision Tree	0,558	0,579	3,288	0,265	0,266	0,006	1
Random Forest	0,454	0,517	3,291	0,403	0,403	6,109	1010,3
Extra Trees	0,488	0,553	3,367	0,357	0,357	1,881	311,1
XGBoost	0,401	0,462	3,358	0,473	0,473	0,485	80,3
LightGBM	0,372	0,427	3,673	0,510	0,510	7,797	1289,4
CatBoost	0,420	0,484	3,719	0,448	0,448	6,195	1024,5
MLP	0,438	0,506	3,656	0,424	0,431	39,178	6529,7
ResNet	0,486	0,539	3,216	0,361	0,369	17,657	2920,1
TabNet	0,588	0,588	3,280	0,226	0,285	17,144	2835,1
SAINT	0,575	0,605	3,135	0,243	0,267	10,400	1719,8
FastFormer	0,490	0,546	3,365	0,355	0,366	16,741	2768,5
FT-Transformer	0,559	0,578	3,340	0,265	0,300	51,990	8597,8
TabTransformer	0,487	0,554	3,267	0,359	0,361	31,962	5285,6
Trivial	0,760	0,684	3,178	0	0	-	-

Tabela 39 – SHAP Médio para a Base de Dados dos Vinhos

Modelo	Álcool (%vol)	Densidade (g/L)	SO ₂ Livre (mg/L)	Acidez Volátil (g/L)	Tinto	Cloreto (g/L)	pH	Açúcar Residual (g/L)	Ácido Cítrico (g/L)	Sulfato (g/L)	SO ₂ Total (mg/L)	Acidez Fixa (g/L)
LASSO	-0,192	-0,272	-0,017	0,100	-0,051	-0,005	0,170	-0,032	-0,002	0,099	0,014	0,044
Ridge	-0,232	-0,287	0,017	0,156	-0,015	0,005	0,169	0,000	-0,004	0,037	-0,029	0,049
Elastic Net	-0,187	-0,268	-0,024	0,109	0,037	0,006	0,155	-0,054	0,004	0,104	0,019	0,034
Relaxed Lasso	-0,211	-0,276	-0,003	0,082	0,054	0,009	0,186	-0,063	0,009	0,085	0,000	0,022
KNN	0,212	-0,072	-0,138	-0,034	0,007	0,055	0,026	-0,153	-0,176	-0,015	-0,061	0,050
SVM	-0,026	0,257	-0,140	0,045	0,070	0,080	0,048	-0,158	-0,033	-0,001	0,051	-0,091
Decision Tree	-0,060	0,384	-0,021	0,174	0,126	0,013	0,008	0,092	0,009	0,035	0,073	-0,003
Random Forest	0,199	0,024	-0,225	0,223	0,004	0,169	0,054	-0,030	-0,013	0,042	-0,010	-0,006
Extra Trees	0,310	0,055	-0,125	0,224	0,086	0,061	0,013	-0,026	-0,039	0,041	-0,008	0,013
XGBoost	0,226	-0,047	-0,206	0,108	0,017	0,076	0,072	0,017	-0,115	-0,014	0,097	0,007
LightGBM	0,197	0,071	-0,184	0,109	0,030	0,124	0,040	-0,027	-0,101	0,048	0,040	-0,028
CatBoost	0,117	0,063	-0,214	0,082	0,032	0,124	0,086	-0,042	-0,077	0,060	0,017	-0,087
MLP	0,026	0,223	-0,020	0,054	0,417	0,044	0,046	-0,060	0,003	0,042	0,043	-0,023
ResNet	0,086	0,128	-0,271	0,069	0,135	0,025	0,029	-0,044	0,129	0,025	0,044	0,015
TabNet	0,281	0,069	0,057	0,028	0,131	0,119	-0,028	0,074	0,066	0,029	0,066	-0,051
SAINT	0,324	0,082	-0,030	0,071	-0,164	0,070	0,016	0,106	0,006	0,046	0,060	0,024
FastFormer	0,089	0,133	-0,122	0,131	0,044	0,178	0,041	0,036	0,080	0,056	0,079	-0,012
FT-Transformer	0,066	0,075	-0,186	0,067	-0,108	0,124	0,097	-0,069	0,060	-0,029	0,046	-0,074
TabTransformer	0,082	0,124	-0,157	0,173	0,267	0,041	0,038	-0,064	0,000	0,004	0,039	0,010
Média	0,069	0,025	-0,106	0,104	0,059	0,069	0,067	-0,026	-0,010	0,037	0,031	-0,006
Desvio Padrão	0,175	0,184	0,094	0,064	0,124	0,055	0,060	0,068	0,070	0,036	0,039	0,043

Tabela 40 – SHAP Absolutos para a Base de Dados dos Vinhos

Modelo	Álcool (%vol)	Densidade (g/L)	Acidez Volátil (g/L)	Açúcar Residual (g/L)	SO ₂ Livre (mg/L)	Tinto	SO ₂ Total (mg/L)	Sulfato (g/L)	pH	Cloreto (g/L)	Acidez Fixa (g/L)	Ácido Cítrico (g/L)
LASSO	0,163	0,175	0,138	0,169	0,052	0,095	0,049	0,054	0,040	0,012	0,049	0,005
Ridge	0,168	0,170	0,140	0,167	0,053	0,093	0,051	0,054	0,039	0,012	0,048	0,006
Elastic Net	0,169	0,169	0,140	0,166	0,054	0,093	0,051	0,054	0,039	0,012	0,047	0,006
Relaxed Lasso	0,163	0,174	0,138	0,169	0,052	0,095	0,049	0,054	0,040	0,012	0,049	0,005
KNN	0,189	0,065	0,110	0,086	0,106	0,019	0,087	0,094	0,077	0,046	0,048	0,074
SVM	0,086	0,215	0,096	0,150	0,087	0,079	0,068	0,037	0,059	0,052	0,047	0,024
Decision Tree	0,138	0,259	0,211	0,143	0,022	0,083	0,059	0,036	0,021	0,014	0,005	0,010
Random Forest	0,339	0,052	0,183	0,048	0,106	0,002	0,062	0,061	0,035	0,073	0,016	0,022
Extra Trees	0,452	0,045	0,217	0,037	0,059	0,040	0,028	0,041	0,022	0,026	0,016	0,018
XGBoost	0,215	0,101	0,145	0,070	0,104	0,007	0,093	0,066	0,047	0,063	0,035	0,055
LightGBM	0,199	0,100	0,149	0,059	0,107	0,014	0,085	0,060	0,045	0,083	0,043	0,056
CatBoost	0,181	0,143	0,111	0,091	0,102	0,012	0,080	0,071	0,058	0,065	0,043	0,041
MLP	0,056	0,189	0,072	0,102	0,100	0,215	0,046	0,052	0,063	0,044	0,037	0,023
ResNet	0,171	0,136	0,117	0,115	0,113	0,041	0,088	0,065	0,042	0,022	0,036	0,053
TabNet	0,281	0,069	0,028	0,074	0,057	0,131	0,066	0,029	0,028	0,119	0,051	0,066
SAINT	0,229	0,059	0,134	0,066	0,056	0,101	0,060	0,078	0,084	0,036	0,029	0,068
FastFormer	0,138	0,125	0,169	0,077	0,070	0,048	0,087	0,081	0,030	0,093	0,035	0,048
FT-Transformer	0,149	0,193	0,102	0,131	0,088	0,039	0,046	0,054	0,056	0,044	0,044	0,055
TabTransformer	0,195	0,118	0,139	0,122	0,095	0,059	0,071	0,057	0,045	0,026	0,023	0,050
Média	0,194	0,135	0,134	0,107	0,078	0,067	0,064	0,058	0,046	0,045	0,037	0,036
Desvio Padrão	0,086	0,059	0,043	0,043	0,026	0,051	0,018	0,016	0,017	0,030	0,013	0,023

Tabela 41 – Feature Permutation para a Base de Dados dos Vinhos

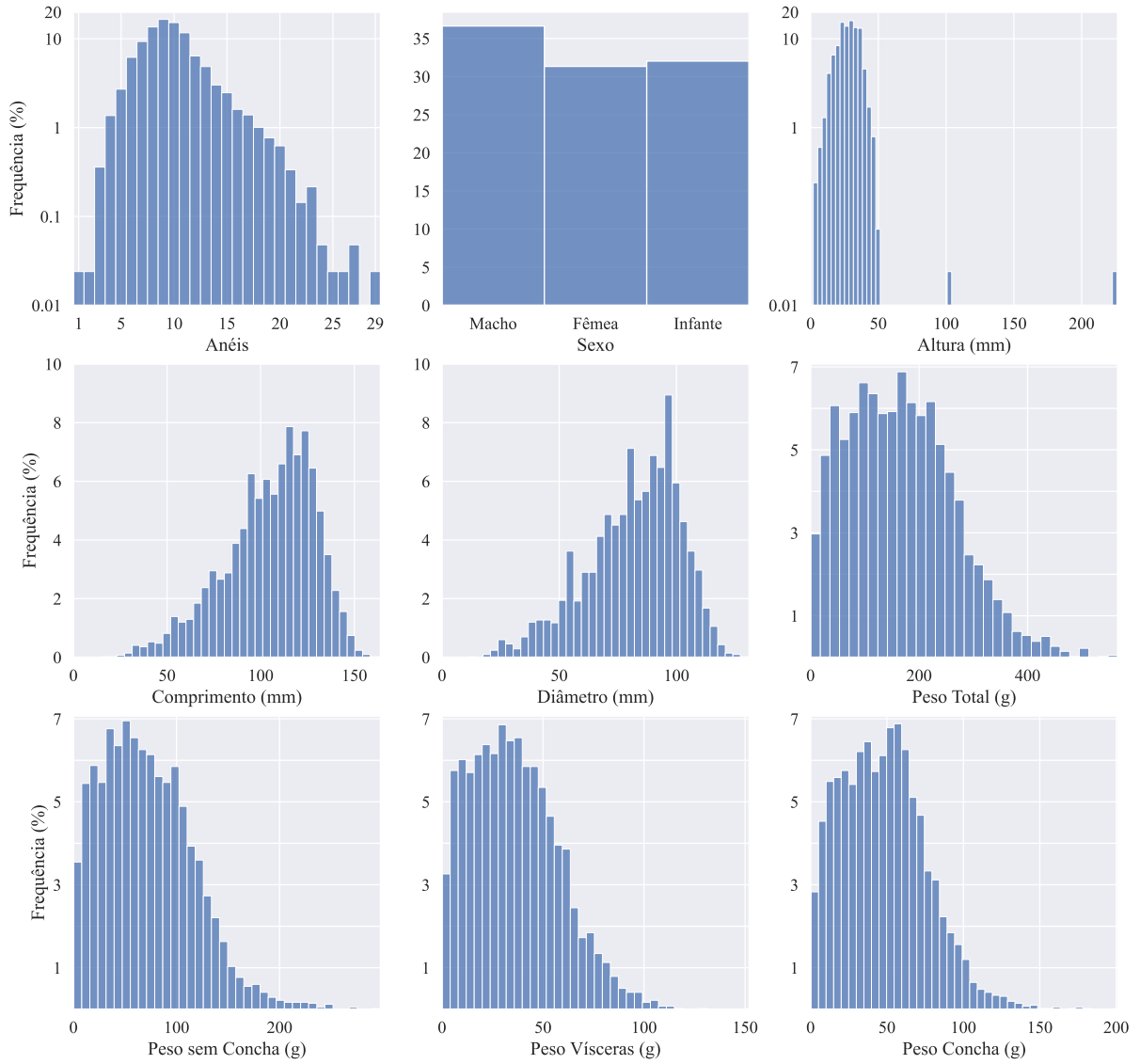
Modelo	Álcool (%vol)	Densidade (g/L)	SO ₂ Livre (mg/L)	Tinto	Açúcar Residual (g/L)	pH	SO ₂ Total (mg/L)	Cloreto (g/L)	Acidez Volátil (g/L)	Sulfato (g/L)	Acidez Fixa (g/L)	Ácido Cítrico (g/L)
LASSO	0,391	0,234	0,038	-0,090	0,068	0,069	0,001	0,021	0,040	0,030	-0,017	0,000
Ridge	0,418	0,225	0,043	-0,079	0,064	0,074	0,000	0,022	0,016	0,042	-0,016	0,001
Elastic Net	0,391	0,215	0,045	-0,091	0,077	0,063	-0,004	0,021	0,038	0,037	-0,018	0,000
Relaxed Lasso	0,399	0,240	0,042	-0,092	0,055	0,069	0,000	0,020	0,028	0,039	-0,013	0,000
KNN	0,188	0,056	0,133	0,017	0,081	0,159	0,116	0,011	0,027	0,182	0,007	-0,025
SVM	0,068	0,195	0,112	0,071	0,098	0,124	0,074	0,083	0,059	0,043	0,040	0,032
Decision Tree	0,568	0	0	0	0,301	0	0	0	0,131	0	0	0
Random Forest	0,398	0,081	0,149	0,002	0,072	0,030	0,072	0,057	0,033	0,039	0,064	-0,004
Extra Trees	0,588	0,048	0,116	0,038	0,041	0,020	0,030	0,033	0,035	0,020	0,017	0,015
XGBoost	0,306	0,056	0,070	0,003	0,062	0,065	0,096	0,084	-0,001	0,157	0,064	-0,037
LightGBM	0,300	0,045	0,084	0,002	0,079	0,087	0,136	0,141	-0,007	0,063	0,050	-0,006
CatBoost	0,192	0,090	0,154	0,016	0,044	0,142	0,127	0,105	0,028	0,090	0,005	-0,007
MLP	0,020	0,256	0,119	0,093	0,076	0,113	0,105	0,019	0,056	0,058	0,047	0,037
ResNet	0,165	0,234	-0,154	0,117	0,009	-0,009	0,046	0,038	0,150	0,052	0,012	0,014
TabNet	0,154	0,261	0,095	0,049	0,055	0,015	0,065	0,147	0,091	-0,017	0,052	-0,002
SAINT	0,159	0,086	0,067	0,025	0,115	-0,058	0,036	0,098	-0,090	0,129	0,067	-0,070
FastFormer	0,045	0,100	0,086	0,292	0,057	-0,067	0,139	0,030	0,051	0,065	0,058	0,011
FT-Transformer	-0,065	-0,097	0,112	-0,325	-0,002	0,013	0,087	0,144	-0,052	-0,016	0,056	0,030
TabTransformer	0,188	0,223	0,013	0,176	0,069	-0,017	-0,014	0,043	0,169	0,018	0,065	0,005
Média	0,257	0,134	0,070	0,012	0,075	0,047	0,058	0,059	0,042	0,054	0,028	-0,000
Desvio Padrão	0,177	0,102	0,068	0,122	0,059	0,062	0,051	0,047	0,061	0,052	0,031	0,024

Tabela 42 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Vinhos

Modelo	Álcool (%vol)	Acidez Volátil (g/L)	Densidade (g/L)	Açúcar Residual (g/L)	Ácido Cítrico (g/L)	Tinto	Sulfato (g/L)	SO ₂ Livre (mg/L)	SO ₂ Total (mg/L)	Acidez Fixa (g/L)	pH	Cloreto (g/L)
LASSO	0,158	-0,145	-0,166	0,164	-0,006	0,089	0,060	0,052	-0,049	0,054	0,040	-0,016
Ridge	0,163	-0,147	-0,161	0,162	-0,006	0,087	0,061	0,053	-0,051	0,053	0,039	-0,017
Elastic Net	0,164	-0,147	-0,160	0,161	-0,006	0,087	0,061	0,053	-0,051	0,053	0,039	-0,017
Relaxed Lasso	0,159	-0,146	-0,166	0,164	-0,006	0,089	0,060	0,052	-0,049	0,054	0,040	-0,016
Decision Tree	0,260	0,296	0,199	0,072	0,002	0,053	0,046	0,014	0,021	0,002	0,018	0,017
Random Forest	0,360	0,161	0,089	0,047	0,045	0,001	0,061	0,076	0,048	0,031	0,033	0,048
Extra Trees	0,504	0,178	0,064	0,030	0,034	0,023	0,032	0,051	0,023	0,019	0,020	0,022
XGBoost	0,220	0,097	0,070	0,061	0,077	0,090	0,066	0,081	0,062	0,053	0,053	0,070
LightGBM	0,091	0,074	0,083	0,077	0,076	0,001	0,075	0,110	0,111	0,128	0,079	0,094
CatBoost	0,134	0,109	0,093	0,095	0,069	0,006	0,089	0,094	0,097	0,064	0,078	0,071
TabNet	0,067	0,111	0,014	0,025	0,526	0,175	0,044	0,002	0,016	0,000	0,021	0,000
Média Absolutos	0,207	0,147	0,115	0,096	0,078	0,064	0,060	0,058	0,053	0,047	0,042	0,035
Desvio Padrão dos Absolutos	0,121	0,056	0,055	0,054	0,145	0,051	0,015	0,030	0,028	0,033	0,020	0,029

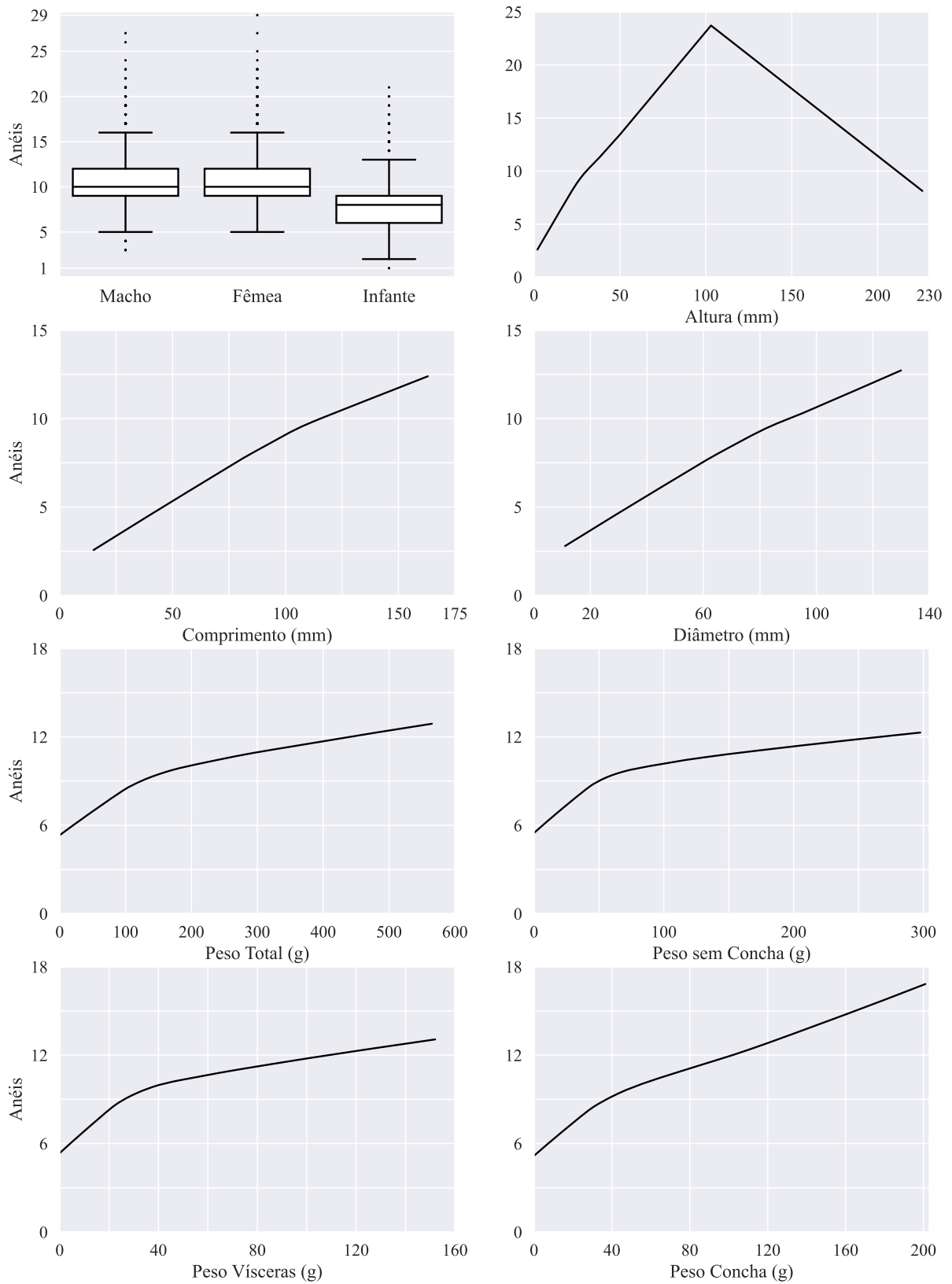
4.9 ANÉIS DE ABALONES

Figura 42 – Histogramas das Variáveis para a Base de Dados dos Abalones



Fonte: Figura do Autor

Figura 43 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados dos Abalones



Fonte: Figura do Autor

Dada a formulação do problema ser similar a um problema de regressão para dados de contagem, devido ao suporte da variável resposta ser não-negativo e inteiro, também foi ajustada um Modelo Linear Generalizado com função de ligação Poisson e penalização Ridge, dado que a média da variável resposta é muito próxima de sua variância, se assemelhando a uma distribuição Poisson em que ambas são iguais. Na tabela 43 é mostrado que apesar dessa semelhança com a distribuição, os resultados obtidos na regressão Poisson não superaram os das regressões lineares penalizadas.

Tabela 43 – Métricas para a Base de Dados dos Abalones

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	5,212	1,625	14,168	0,488	0,488	0,042	8,8
Ridge	5,217	1,625	14,226	0,487	0,488	0,005	1
Poisson	6,549	1,683	29,216	0,356	0,357	0,040	8,4
Elastic Net	5,210	1,625	14,177	0,488	0,488	0,040	8,4
Relaxed Lasso	5,213	1,625	14,167	0,488	0,488	0,050	10,6
KNN	5,240	1,579	10,961	0,485	0,488	0,009	1,8
SVM	4,717	1,480	10,608	0,536	0,542	0,847	178,5
Decision Tree	4,931	1,561	11,381	0,515	0,516	0,010	2,0
Random Forest	4,670	1,506	10,779	0,541	0,541	2,927	616,5
Extra Trees	4,679	1,512	10,705	0,540	0,540	1,533	322,9
XGBoost	5,388	1,651	14,281	0,471	0,471	0,034	7,2
LightGBM	4,729	1,530	10,248	0,535	0,535	0,763	160,8
CatBoost	4,865	1,568	10,229	0,522	0,522	5,734	1207,9
MLP	4,536	1,499	10,004	0,554	0,555	8,899	1874,6
ResNet	4,840	1,566	9,931	0,524	0,526	30,052	6330,3
TabNet	4,717	1,548	9,747	0,536	0,538	28,152	5930,1
SAINT	5,494	1,600	10,677	0,460	0,508	34,751	7320,0
FastFormer	5,538	1,637	11,141	0,456	0,484	13,662	2877,8
FT-Transformer	4,859	1,507	11,030	0,522	0,527	25,367	5343,4
TabTransformer	5,044	1,589	10,154	0,504	0,505	27,693	5833,4
Trivial	10,176	2,348	13,112	0	0	-	-

Originalmente, a variável que tratava do sexo de cada animal possuía as categorias Macho, Fêmea e Infante, devido ao fato que só é possível diferenciar machos de fêmeas, visualmente, após esses gastrópodes atingirem a maturidade. Após o *One-Hot Encoding*, uma dada amostra se refere a um macho quando ambas as variáveis binárias Fêmea e Infante são zero, sendo que elas não podem assumir simultaneamente o valor 1 para um único animal.

Tabela 44 – SHAP Médio para a Base de Dados dos Abalones

Modelo	Peso Sem Concha	Peso Total	Peso Concha	Diâmetro	Peso Vísceras	Infante	Comprimento	Altura	Fêmea
LASSO	-0,382	0,223	-0,065	0,085	-0,112	0,008	0	0,063	0,062
Ridge	-0,343	0,192	0,012	0,128	-0,156	-0,049	-0,015	0,033	0,072
Poisson	-0,280	0,283	0,094	0,103	-0,085	0,037	0,035	0,083	-0,001
Elastic Net	-0,347	0,303	-0,013	0,064	-0,111	-0,058	0	0,070	-0,033
Relaxed Lasso	-0,453	0,232	-0,048	0,145	-0,086	-0,028	0	0,009	0
KNN	0,152	0,036	0,308	-0,093	-0,022	-0,086	-0,041	0,114	0,149
SVM	0,264	0,099	-0,155	-0,171	-0,104	0,019	-0,138	0,000	0,049
Decision Tree	0,045	-0,001	0,451	-0,128	-0,004	-0,144	-0,002	0,225	0,000
Random Forest	0,107	-0,129	-0,015	-0,179	-0,103	-0,274	-0,168	0,010	0,016
Extra Trees	-0,044	-0,074	0,164	-0,262	-0,100	-0,098	-0,212	-0,046	-0,001
XGBoost	0,094	-0,069	-0,137	-0,014	0,033	0,312	-0,029	-0,213	0,100
LightGBM	-0,039	0,070	-0,244	-0,243	-0,052	-0,105	-0,159	0,055	0,032
CatBoost	-0,092	0,237	-0,004	-0,305	-0,168	-0,017	-0,056	0,044	0,077
MLP	0,213	-0,369	-0,097	-0,056	-0,020	-0,121	-0,047	-0,031	-0,047
ResNet	0,335	-0,228	-0,040	-0,125	-0,025	0,116	-0,069	0,026	0,036
TabNet	-0,139	0,291	0,103	0,011	-0,151	-0,121	-0,131	0,044	-0,010
SAINT	0,228	0,043	-0,154	-0,030	-0,183	-0,034	-0,131	0,077	-0,120
FastFormer	0,202	0,065	0,128	-0,102	0,063	0,102	-0,095	-0,146	-0,096
FT-Transformer	0,391	0,060	-0,231	-0,036	-0,073	0,020	-0,175	-0,008	0,005
TabTransformer	0,305	-0,093	-0,060	-0,217	-0,125	-0,003	-0,188	-0,006	0,004
Média	0,011	0,059	-0,000	-0,071	-0,079	-0,026	-0,081	0,020	0,015
Desvio Padrão	0,256	0,178	0,168	0,131	0,065	0,116	0,074	0,088	0,061

Tabela 45 – SHAP Absolutos para a Base de Dados dos Abalones

Modelo	Peso Sem Concha	Peso Concha	Peso Total	Peso Vísceras	Diâmetro	Altura	Comprimento	Infante	Fêmea
LASSO	0,336	0,103	0,322	0,089	0,067	0,046	0	0,037	0,000
Ridge	0,335	0,103	0,320	0,091	0,066	0,046	0,001	0,038	0,000
Poisson	0,327	0,079	0,272	0,085	0,097	0,056	0,034	0,049	0,001
Elastic Net	0,336	0,100	0,325	0,090	0,067	0,045	0	0,037	0
Relaxed Lasso	0,337	0,105	0,319	0,088	0,067	0,046	0	0,038	0
KNN	0,276	0,303	0,068	0,055	0,056	0,119	0,047	0,033	0,041
SVM	0,325	0,204	0,300	0,036	0,037	0,029	0,032	0,023	0,013
Decision Tree	0,320	0,539	0,010	0,007	0,032	0,056	0,002	0,034	0
Random Forest	0,320	0,496	0,076	0,012	0,021	0,019	0,023	0,030	0,002
Extra Trees	0,246	0,450	0,054	0,021	0,052	0,053	0,049	0,060	0,015
XGBoost	0,334	0,227	0,180	0,068	0,007	0,059	0,074	0,051	0,000
LightGBM	0,308	0,250	0,235	0,037	0,065	0,033	0,040	0,023	0,008
CatBoost	0,313	0,137	0,277	0,070	0,095	0,035	0,044	0,008	0,020
MLP	0,315	0,162	0,341	0,051	0,009	0,033	0,061	0,021	0,007
ResNet	0,313	0,199	0,317	0,024	0,030	0,022	0,055	0,024	0,015
TabNet	0,139	0,103	0,291	0,151	0,011	0,044	0,131	0,121	0,010
SAINT	0,263	0,365	0,095	0,075	0,037	0,049	0,051	0,022	0,044
FastFormer	0,277	0,131	0,236	0,039	0,064	0,070	0,055	0,082	0,046
FT-Transformer	0,334	0,370	0,167	0,023	0,032	0,007	0,047	0,014	0,005
TabTransformer	0,336	0,262	0,254	0,018	0,076	0,006	0,036	0,006	0,006
Média	0,305	0,234	0,223	0,057	0,049	0,044	0,039	0,038	0,012
Desvio Padrão	0,046	0,139	0,105	0,035	0,026	0,024	0,031	0,026	0,015

Tabela 46 – Feature Permutation para a Base de Dados dos Abalones

Modelo	Peso Total	Peso Sem Concha	Peso Concha	Peso Visceras	Diâmetro	Altura	Comprimento	Infante	Fêmea
LASSO	0,423	0,025	0,120	-0,225	0,081	0,077	0,000	0,049	-0,000
Ridge	0,425	0,011	0,114	-0,234	0,081	0,080	0,002	0,051	-0,001
Poisson	0,261	-0,091	0,084	-0,251	0,104	0,083	0,052	0,071	-0,004
Elastic Net	0,432	0,022	0,120	-0,230	0,080	0,070	0,000	0,045	-0,001
Relaxed Lasso	0,370	0,108	0,108	-0,220	0,072	0,072	0,000	0,049	0,000
KNN	-0,120	0,080	-0,110	-0,390	-0,000	-0,134	-0,153	-0,011	-0,001
SVM	0,448	-0,142	0,044	-0,303	0,015	0,010	-0,010	0,004	-0,023
Decision Tree	0,000	0,296	0,432	0,000	-0,272	0,000	0,000	0,000	0,000
Random Forest	0,218	0,181	0,438	-0,031	-0,079	0,020	-0,030	0,002	-0,000
Extra Trees	0,137	0,206	0,427	-0,083	-0,032	0,019	-0,034	0,051	-0,011
XGBoost	0,210	0,099	0,248	-0,167	0,007	0,094	0,109	0,065	0,000
LightGBM	0,314	0,039	0,246	-0,252	-0,110	-0,019	-0,002	0,004	-0,014
CatBoost	0,347	0,220	0,118	-0,182	-0,043	-0,007	-0,018	0,010	0,055
MLP	0,454	0,079	0,155	-0,156	-0,003	0,064	-0,072	-0,012	-0,004
ResNet	-0,351	-0,331	-0,184	-0,030	-0,026	-0,005	-0,067	-0,004	-0,003
TabNet	-0,151	-0,421	0,078	-0,068	0,125	0,097	0,025	0,027	0,007
SAINT	0,028	-0,528	-0,213	-0,080	-0,041	0,004	-0,061	-0,021	-0,025
FastFormer	-0,317	-0,417	0,041	0,012	0,000	-0,041	0,067	-0,094	-0,009
FT-Transformer	-0,075	-0,424	-0,450	-0,003	-0,012	-0,000	-0,033	-0,002	0,001
TabTransformer	-0,177	-0,454	-0,226	-0,072	-0,030	-0,000	-0,042	-0,000	0,000
Média	0,144	-0,072	0,080	-0,148	-0,004	0,024	-0,013	0,014	-0,002
Desvio Padrão	0,262	0,255	0,225	0,111	0,087	0,055	0,054	0,037	0,015

Tabela 47 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados dos Abalones

Modelo	Peso Concha	Peso Sem Concha	Peso Total	Altura	Peso Visceras	Diâmetro	Infante	Comprimento	Fêmea
LASSO	0,103	-0,337	0,315	0,061	-0,087	0,065	-0,032	0	-0,000
Ridge	0,103	-0,336	0,314	0,061	-0,089	0,064	-0,032	0,001	-0,000
Poisson	0,077	-0,328	0,267	0,072	-0,082	0,096	-0,043	0,034	-0,001
Elastic Net	0,100	-0,337	0,318	0,060	-0,088	0,065	-0,032	0	-0,000
Relaxed Lasso	0,105	-0,337	0,312	0,061	-0,086	0,065	-0,032	0	0
Decision Tree	0,790	0,146	0,006	0,010	0,003	0,019	0,025	0,001	0
Random Forest	0,703	0,162	0,037	0,016	0,015	0,029	0,023	0,012	0,003
Extra Trees	0,330	0,114	0,097	0,099	0,061	0,105	0,107	0,076	0,010
XGBoost	0,225	-0,333	0,174	0,079	-0,066	0,005	-0,043	0,074	0,000
LightGBM	0,183	0,182	0,148	0,090	0,159	0,098	0,013	0,110	0,018
CatBoost	0,173	0,185	0,142	0,108	0,135	0,093	0,028	0,093	0,042
TabNet	0,146	0,119	0,011	0,288	0,036	0,057	0,277	0,043	0,022
Média Absolutos	0,256	0,243	0,175	0,084	0,074	0,067	0,058	0,034	0,008
Desvio Padrão dos Absolutos	0,231	0,094	0,119	0,068	0,043	0,026	0,070	0,038	0,013

4.10 CARDIOTOCOGRAFIA

Figura 44 – Histograma das Variáveis para a Base de Dados da Cardiotocografia



Fonte: Figura do Autor

Tabela 48 – Métricas para a Base de Dados da Cardiotocografia

Modelo	Acurácia	F0,5	F1	F2	AUROC	Recall	Precisão	Log Loss	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,912	0,906	0,907	0,910	0,973	0,912	0,906	0,218	0,752	0,755	1,804	1217
Ridge	0,903	0,896	0,897	0,900	0,971	0,903	0,896	0,226	0,725	0,728	0,693	467
Elastic Net	0,906	0,899	0,900	0,903	0,973	0,906	0,899	0,221	0,733	0,736	2,120	1430
KNN	0,884	0,870	0,871	0,878	0,954	0,884	0,873	0,450	0,645	0,660	0,001	1
SVM	0,918	0,916	0,916	0,917	0,979	0,918	0,917	0,205	0,776	0,777	0,405	273
Decision Tree	0,878	0,871	0,872	0,875	0,936	0,878	0,872	0,590	0,639	0,645	0,007	4
Random Forest	0,925	0,925	0,923	0,924	0,985	0,925	0,927	0,173	0,795	0,796	2,387	1610
Extra Trees	0,931	0,929	0,927	0,929	0,987	0,931	0,932	0,194	0,803	0,808	1,584	1068
XGBoost	0,940	0,942	0,941	0,941	0,989	0,940	0,943	0,169	0,843	0,844	0,156	105
LightGBM	0,944	0,945	0,944	0,943	0,988	0,944	0,946	0,277	0,850	0,851	1,043	703
CatBoost	0,953	0,953	0,953	0,953	0,988	0,953	0,954	0,161	0,872	0,873	105,316	71030
MLP	0,906	0,902	0,903	0,904	0,973	0,906	0,902	0,233	0,735	0,737	0,597	403
ResNet	0,906	0,913	0,909	0,907	0,973	0,906	0,915	0,235	0,759	0,761	23,389	15775
TabNet	0,790	0,837	0,810	0,794	0,893	0,790	0,859	0,622	0,544	0,570	7,540	5085
SAINT	0,887	0,899	0,893	0,889	0,975	0,887	0,904	0,261	0,718	0,723	8,948	6035
FastFormer	0,818	0,854	0,834	0,822	0,956	0,818	0,871	0,413	0,590	0,610	7,492	5053
FT-Transformer	0,890	0,904	0,897	0,892	0,974	0,890	0,910	0,249	0,730	0,737	11,333	7643
TabTransformer	0,881	0,893	0,887	0,883	0,962	0,881	0,899	0,318	0,706	0,712	12,262	8270
Trivial	0,777	0,633	0,680	0,735	0,5	0,777	0,604	0,678	0,735	0,737	-	-

É possível notar na tabela 48 que a escolha, feita durante a seleção de hiperparâmetros, de utilizar o *Ordered Boosting* no CatBoost ocasionou um expressivo aumento do custo computacional do treinamento do modelo final, acompanhado da obtenção das melhores métricas.

Foi escolhido representar a importância das variáveis para somente uma das classes considerada mais importante, aquela que correspondem aos exames classificados pelos especialistas como patológicos.

Tabela 49 – SHAP Médio para a Base de Dados da Cardiotocografia

Modelo	Desacelerações Prolongadas (1/s)	% Tempo Variabilidade Anormal de Curto Prazo	Média Histograma	% Tempo Variabilidade Anormal de Longo Prazo	Nº de Acelerações (1/s)	Mediana Histograma	Variância Histograma	Moda Histograma	Média na Variabilidade de Curto Prazo	Contrações Uterinas (1/s)
LASSO	0,126	0,210	0,069	0,103	0,176	0,118	0,088	0,054	0,003	0,006
Ridge	0,155	0,204	0,073	0,112	0,103	0,101	0,086	0,072	0,004	0,006
Elastic Net	0,147	0,214	0,090	0,111	0,108	0,105	0,089	0,060	0,007	0,005
KNN	0,272	0,056	0,089	0,177	0,008	0,063	0,039	0,076	0,028	0,059
SVM	0,147	0,092	0,090	0,147	0,029	0,076	0,061	0,043	-0,040	0,071
Decision Tree	0,000	0,345	0,239	0,250	0,015	0,004	0,001	-0,001	-0,000	-0,039
Random Forest	0,115	0,168	0,260	0,149	0,053	0,022	0,025	0,042	0,110	0,026
Extra Trees	0,132	0,130	0,169	0,142	0,079	0,079	0,021	0,070	0,061	0,050
XGBoost	0,131	0,122	0,165	0,086	0,058	0,065	0,086	0,070	0,039	0,029
LightGBM	0,105	0,062	0,186	0,165	0,111	0,076	0,036	0,059	0,039	0,036
CatBoost	0,101	0,163	0,107	0,142	0,198	0,033	0,037	0,042	0,049	0,032
MLP	0,155	0,158	0,077	0,120	0,070	0,050	0,080	0,063	-0,031	0,019
ResNet	-0,106	-0,031	0,196	-0,079	0,054	0,000	0,079	0,011	-0,044	0,043
SAINT	-0,241	-0,069	0,073	0,029	0,028	0,075	0,015	-0,038	-0,063	0,036
FastFormer	-0,042	-0,016	0,126	0,102	0,231	-0,005	0,037	-0,005	0,011	0,017
FT-Transformer	-0,338	-0,022	0,075	0,052	0,087	-0,001	0,044	-0,044	-0,090	-0,052
TabTransformer	0,239	0,053	0,011	-0,033	0,104	0,144	0,006	-0,038	0,009	0,082
Média	0,065	0,108	0,123	0,104	0,089	0,059	0,049	0,032	0,005	0,025
Desvio Padrão	0,157	0,105	0,066	0,076	0,062	0,043	0,030	0,041	0,047	0,033

Tabela 50 – SHAP Absolutos para a Base de Dados da Cardiocografia

Modelo	% Tempo Variabilidade Anormal de Curto Prazo	% Tempo Variabilidade Anormal de Longo Prazo	Desacelerações Pro-longadas (1/s)	Média Histograma	Nº de Acelerações (1/s)	Mediana Histograma	Moda Histograma	Contrações Uterinas (1/s)	Média na Variabilidade de Curto Prazo	Variância Histograma
LASSO	0,146	0,082	0,084	0,080	0,093	0,105	0,045	0,029	0,021	0,095
Ridge	0,130	0,083	0,096	0,078	0,052	0,084	0,056	0,030	0,041	0,085
Elastic Net	0,137	0,083	0,092	0,093	0,055	0,088	0,048	0,030	0,039	0,090
KNN	0,080	0,154	0,233	0,078	0,010	0,057	0,068	0,066	0,026	0,035
SVM	0,091	0,129	0,123	0,073	0,032	0,080	0,042	0,075	0,048	0,054
Decision Tree	0,332	0,242	0,001	0,220	0,014	0,006	0,003	0,037	0,001	0,001
Random Forest	0,172	0,147	0,108	0,246	0,050	0,022	0,042	0,031	0,105	0,024
Extra Trees	0,132	0,131	0,176	0,148	0,078	0,070	0,063	0,051	0,054	0,020
XGBoost	0,115	0,082	0,108	0,137	0,061	0,056	0,063	0,043	0,041	0,073
LightGBM	0,076	0,151	0,095	0,167	0,100	0,068	0,053	0,054	0,037	0,035
CatBoost	0,152	0,132	0,094	0,100	0,184	0,031	0,040	0,049	0,045	0,037
MLP	0,141	0,120	0,128	0,064	0,054	0,040	0,063	0,038	0,019	0,109
ResNet	0,067	0,092	0,052	0,066	0,101	0,065	0,083	0,065	0,080	0,035
SAINT	0,051	0,031	0,164	0,082	0,076	0,052	0,051	0,068	0,065	0,016
FastFormer	0,042	0,088	0,042	0,097	0,186	0,024	0,024	0,034	0,060	0,034
FT-Transformer	0,055	0,075	0,170	0,047	0,103	0,052	0,070	0,068	0,071	0,028
TabTransformer	0,051	0,061	0,103	0,088	0,073	0,066	0,040	0,055	0,061	0,025
Média	0,116	0,111	0,110	0,110	0,078	0,057	0,050	0,048	0,048	0,047
Desvio Padrão	0,067	0,047	0,053	0,054	0,048	0,025	0,018	0,015	0,024	0,031

Tabela 51 – Feature Permutation para a Base de Dados da Cardiocografia

Modelo	Desacelerações Pro-longadas (1/s)	Média Histograma	Variância Histograma	Nº de Acelerações (1/s)	% Tempo Variabilidade Anormal de Longo Prazo	% Tempo Variabilidade Anormal de Curto Prazo	Média na Variabilidade de Curto Prazo	Moda Histograma	Desacelerações Leves (1/s)	Contrações Uterinas (1/s)
LASSO	-0,096	0,000	-0,317	-0,500	-0,009	-0,054	-0,001	0,000	-0,000	-0,000
Ridge	-0,321	0,000	-0,337	-0,160	-0,026	-0,075	-0,007	-0,000	-0,002	-0,001
Elastic Net	-0,231	0,000	-0,549	-0,111	-0,013	-0,056	-0,003	-0,000	-0,001	-0,000
KNN	-0,621	0	-0,015	0	-0,256	0	0	0	-0,041	0
SVM	-0,175	-0,010	-0,317	-0,002	-0,089	-0,020	-0,003	-0,069	-0,005	-0,004
Decision Tree	0	-0,826	0	-0,037	0	-0,137	0	0	0	0
Random Forest	-0,228	-0,484	-0,023	0	-0,001	-0,145	-0,009	-0,091	-0,001	-0,001
Extra Trees	-0,365	-0,216	-0,032	-0,012	-0,039	-0,107	-0,028	-0,118	-0,004	-0,001
XGBoost	-0,189	-0,161	-0,038	-0,190	-0,143	-0,203	0,004	0,004	0,004	0,005
LightGBM	-0,012	-0,186	-0,000	-0,000	-0,715	-0,086	-0,000	-0,000	0,000	-0,000
CatBoost	-0,012	-0,007	-0,001	-0,805	-0,049	-0,107	-0,002	-0,003	-0,001	0,000
MLP	-0,471	0,001	-0,301	-0,043	-0,109	-0,036	-0,003	0,001	0,001	0,001
ResNet	0,064	0,174	0,063	-0,068	0,109	-0,034	0,105	0,085	0,024	0,082
SAINT	0,037	0,188	0,050	0,037	0,163	0,038	-0,059	0,031	0,129	-0,034
FastFormer	0,123	0,149	0,131	0,068	0,057	0,054	0,066	-0,015	0,023	0,038
FT-Transformer	0,072	0,096	0,131	0,013	0,035	0,088	0,078	0,009	0,152	-0,049
TabTransformer	0,132	0,032	-0,021	0,203	0,163	0,061	0,112	-0,025	-0,032	0,053
Média	-0,135	-0,074	-0,093	-0,095	-0,054	-0,048	0,015	-0,011	0,015	0,005
Desvio Padrão	0,211	0,246	0,189	0,227	0,194	0,078	0,045	0,045	0,049	0,029

Tabela 52 – Importância das Variáveis Internamente nos Modelos Lineares para a Base de Dados da Cardiocografia

Modelo	Frequência Cardíaca Fetal Base	% Tempo Variabilidade Anormal de Curto Prazo	Mediana Histograma	Variância Histograma	Nº Picos Histograma	% Tempo Variabilidade Anormal de Longo Prazo	Moda Histograma	Média Histograma	Desacelerações Leves (1/s)	Movimentos Fetais (1/s)
LASSO	0,159	0,149	-0,154	0,117	-0,092	0,092	-0,077	-0,004	0,032	0,037
Ridge	0,106	0,115	-0,079	0,075	-0,068	0,066	-0,056	-0,086	0,034	0,030
Elastic Net	0,148	0,119	-0,112	0,094	-0,086	0,082	-0,074	-0,062	0,037	0,029
Média	0,137	0,128	-0,115	0,096	-0,082	0,080	-0,069	-0,050	0,034	0,032
Desvio Padrão	0,023	0,015	0,031	0,017	0,010	0,011	0,009	0,034	0,002	0,004

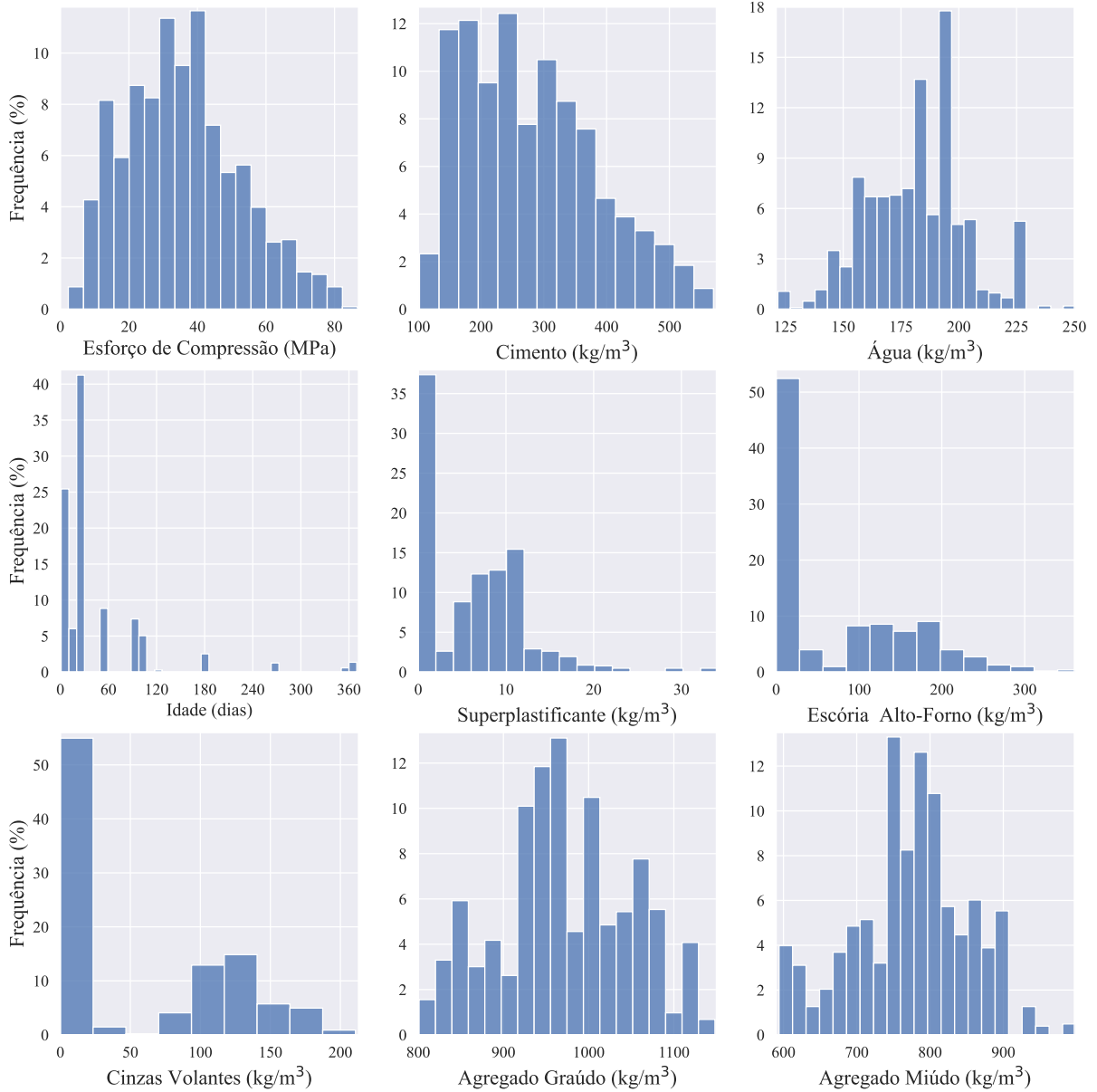
A importância interna para os métodos baseados em Árvores de Decisão e TabNet são mais gerais, não dependendo da classe escolhida.

Tabela 53 – Importância das Variáveis Internamente para a Base de Dados da Cardiotocografia

Modelo	% Tempo Variabilidade Anormal de Curto Prazo	% Tempo Variabilidade Anormal de Longo Prazo	Média Histograma	Nº de Acelerações (1/s)	Moda Histograma	Média na Variabilidade de Curto Prazo	Desacelerações Prolongadas (1/s)	Variância Histograma	Contrações Uterinas (1/s)	Mediana Histograma
Decision Tree	0,369	0,230	0,186	0,053	0,018	0,000	0,000	0,000	0,020	0,058
Random Forest	0,251	0,143	0,151	0,048	0,032	0,118	0,056	0,016	0,032	0,034
Extra Trees	0,183	0,160	0,100	0,096	0,048	0,073	0,099	0,017	0,055	0,048
XGBoost	0,064	0,077	0,125	0,079	0,042	0,117	0,170	0,083	0,033	0,021
LightGBM	0,109	0,093	0,077	0,023	0,057	0,037	0,012	0,046	0,052	0,062
CatBoost	0,142	0,077	0,051	0,125	0,047	0,046	0,019	0,051	0,071	0,039
TabNet	0,137	0,186	0,001	0,127	0,218	0,020	0,013	0,109	0,046	0,043
Média	0,179	0,138	0,099	0,079	0,066	0,059	0,053	0,046	0,044	0,044
Desvio Padrão	0,094	0,054	0,058	0,037	0,063	0,043	0,057	0,036	0,016	0,013

4.11 RESISTÊNCIA À COMPRESSÃO DO CONCRETO

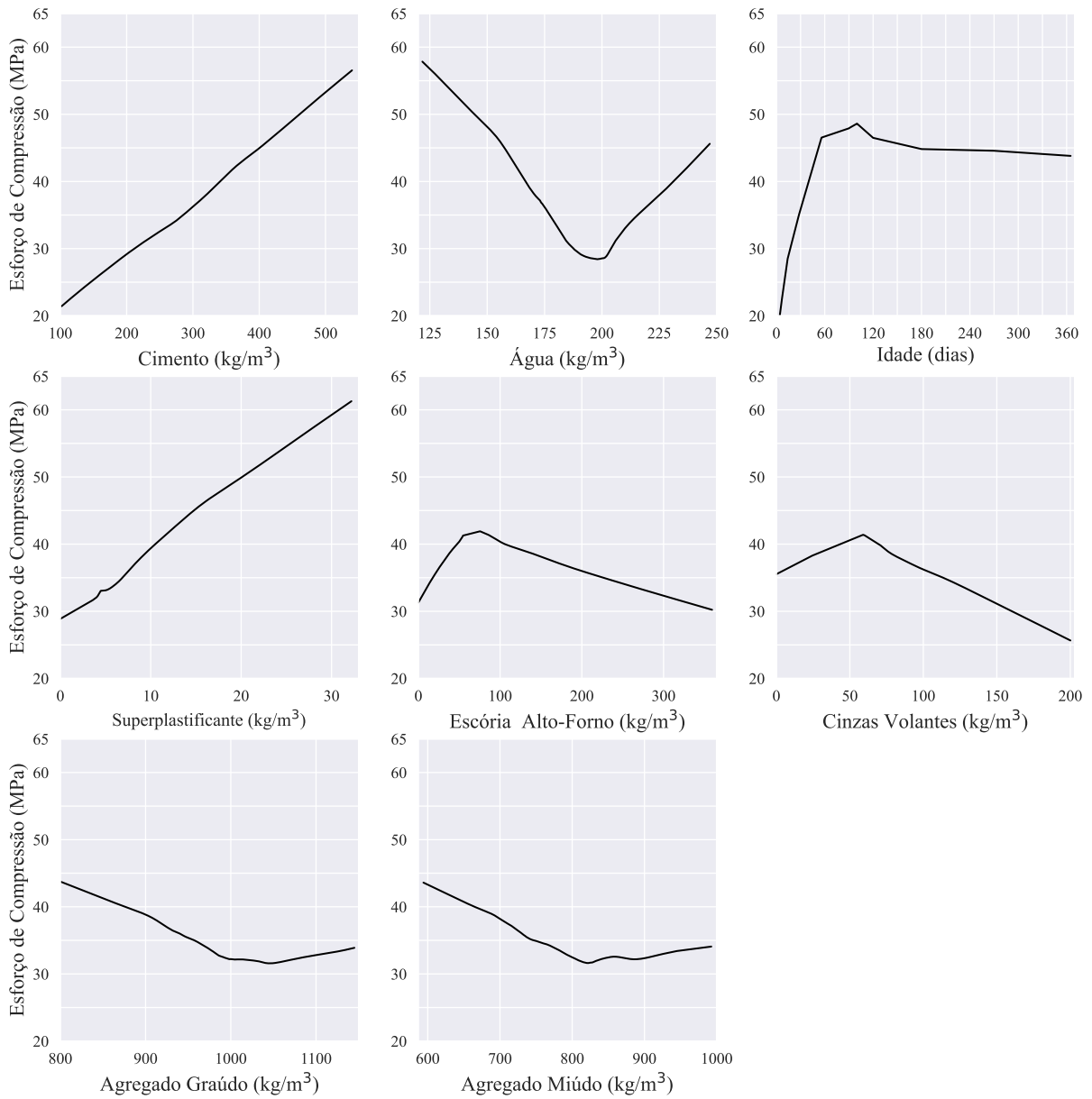
Figura 45 – Histogramas das Variáveis para a Base de Dados do Concreto



Fonte: Figura do Autor

A variável Idade trata da quantidade de dias que se passaram entre a fabricação do concreto e o teste de resistência à compressão, todas as outras variáveis são relacionadas à concentração dos aditivos e estão na unidade de kg adicionados por m³ de concreto.

Figura 46 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados do Concreto



Fonte: Figura do Autor

Tabela 54 – Métricas para a Base de Dados do Concreto

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	96,404	7,887	25,536	0,624	0,624	0,003	1,3
Ridge	96,480	7,893	25,531	0,623	0,623	0,006	3,0
Elastic Net	95,532	7,976	24,680	0,627	0,627	0,002	1,1
Relaxed Lasso	95,146	7,901	25,142	0,628	0,629	0,003	1,2
KNN	55,729	5,679	29,016	0,782	0,792	0,002	1
SVM	34,505	4,308	24,821	0,865	0,870	0,421	192,8
Decision Tree	79,688	6,498	29,694	0,689	0,699	0,002	1,1
Random Forest	27,148	4,089	17,320	0,894	0,895	1,371	627,6
Extra Trees	31,428	4,494	15,968	0,877	0,878	0,978	447,5
XGBoost	25,646	3,841	14,984	0,900	0,902	0,094	42,9
LightGBM	17,588	3,157	14,369	0,931	0,931	0,537	245,8
CatBoost	14,512	2,782	14,041	0,943	0,943	6,319	2891,3
MLP	22,376	3,575	18,349	0,913	0,922	31,953	14621,0
ResNet	24,599	3,656	16,406	0,904	0,904	38,293	17522,1
TabNet	50,450	5,474	25,159	0,803	0,813	13,184	6032,7
SAINT	93,049	6,942	36,095	0,637	0,646	15,306	7003,8
FastFormer	55,879	5,954	17,551	0,782	0,873	11,570	5294,0
FT-Transformer	62,867	6,070	20,527	0,754	0,842	5,867	2684,8
TabTransformer	55,483	5,752	28,394	0,783	0,788	6,650	3042,8
Trivial	256,075	12,996	43,718	0	0	-	-

Tabela 55 – SHAP Médio para a Base de Dados do Concreto

Modelo	Idade (dias)	Cimento (kg/m ³)	Superplas-tificante (kg/m ³)	Escória Alto-Forno (kg/m ³)	Água (kg/m ³)	Agregado Graúdo (kg/m ³)	Agregado Miúdo (kg/m ³)	Cinzas Volantes (kg/m ³)
LASSO	0,017	0,411	0,012	-0,007	-0,227	-0,265	0,022	-0,038
Ridge	-0,025	0,403	-0,064	0,011	-0,182	-0,163	-0,118	0,033
Elastic Net	-0,214	0,292	0,026	0,134	-0,159	0,050	0,087	-0,036
Relaxed Lasso	-0,035	0,502	-0,106	-0,038	0,156	0	0	-0,162
KNN	-0,224	-0,140	-0,234	-0,161	0,114	0,102	-0,019	-0,007
SVM	-0,538	-0,150	-0,098	-0,062	-0,011	0,063	-0,005	0,073
Decision Tree	-0,579	0,000	0,144	-0,002	0,126	-0,002	0,109	0,037
Random Forest	-0,772	-0,014	0,028	-0,137	0,019	0,016	0,002	0,011
Extra Trees	-0,569	-0,090	-0,155	-0,080	0,070	0,019	0,002	0,014
XGBoost	-0,576	-0,070	-0,099	-0,117	0,029	0,063	-0,023	0,023
LightGBM	-0,433	-0,229	-0,097	-0,077	-0,001	0,106	0,029	-0,028
CatBoost	-0,555	-0,130	-0,039	-0,146	0,071	0,026	0,014	0,019
MLP	-0,561	-0,158	-0,083	0,005	0,002	0,048	-0,098	0,045
ResNet	-0,099	-0,091	-0,181	-0,346	0,052	0,069	-0,113	0,050
TabNet	-0,094	0,315	0,138	-0,040	0,047	-0,037	0,198	0,132
SAINT	0,124	-0,199	-0,310	-0,152	-0,022	-0,016	-0,069	-0,107
FastFormer	-0,391	-0,082	-0,132	-0,251	0,040	0,043	-0,060	-0,000
FT-Transformer	-0,216	-0,078	-0,327	-0,209	0,088	0,047	-0,017	0,019
TabTransformer	-0,542	0,096	-0,032	-0,097	0,122	0,043	0,023	0,045
Média	-0,331	0,031	-0,085	-0,093	0,017	0,011	-0,002	0,007
Desvio Padrão	0,255	0,226	0,124	0,105	0,102	0,087	0,075	0,062

Tabela 56 – SHAP Absolutos para a Base de Dados do Concreto

Modelo	Cimento (kg/m ³)	Idade (dias)	Escória Alto-Forno (kg/m ³)	Água (kg/m ³)	Superplas- tificante (kg/m ³)	Cinzas Volantes (kg/m ³)	Agregado Miúdo (kg/m ³)	Agregado Graúdo (kg/m ³)
LASSO	0,302	0,131	0,228	0,085	0,041	0,156	0,028	0,028
Ridge	0,301	0,131	0,228	0,084	0,044	0,154	0,028	0,029
Elastic Net	0,295	0,154	0,203	0,141	0,062	0,114	0,022	0,009
Relaxed Lasso	0,313	0,151	0,224	0,128	0,045	0,140	0	0
KNN	0,206	0,181	0,137	0,125	0,142	0,056	0,091	0,063
SVM	0,248	0,250	0,194	0,061	0,044	0,109	0,046	0,048
Decision Tree	0,189	0,414	0,017	0,117	0,135	0,024	0,102	0,001
Random Forest	0,225	0,365	0,112	0,116	0,108	0,014	0,046	0,014
Extra Trees	0,236	0,332	0,117	0,080	0,165	0,027	0,028	0,016
XGBoost	0,181	0,362	0,116	0,120	0,100	0,026	0,052	0,043
LightGBM	0,200	0,285	0,120	0,164	0,075	0,020	0,071	0,065
CatBoost	0,221	0,332	0,158	0,123	0,081	0,020	0,040	0,026
MLP	0,254	0,213	0,193	0,055	0,033	0,158	0,049	0,045
ResNet	0,297	0,156	0,221	0,038	0,043	0,132	0,056	0,056
TabNet	0,315	0,094	0,040	0,047	0,138	0,132	0,198	0,037
SAINT	0,265	0,119	0,097	0,164	0,158	0,061	0,095	0,040
FastFormer	0,229	0,287	0,179	0,110	0,066	0,010	0,064	0,055
FT-Transformer	0,277	0,212	0,221	0,131	0,083	0,050	0,012	0,014
TabTransformer	0,239	0,244	0,120	0,183	0,035	0,032	0,082	0,063
Média	0,252	0,232	0,154	0,109	0,084	0,076	0,059	0,034
Desvio Padrão	0,042	0,094	0,062	0,040	0,043	0,055	0,043	0,021

Tabela 57 – Feature Permutation para a Base de Dados do Concreto

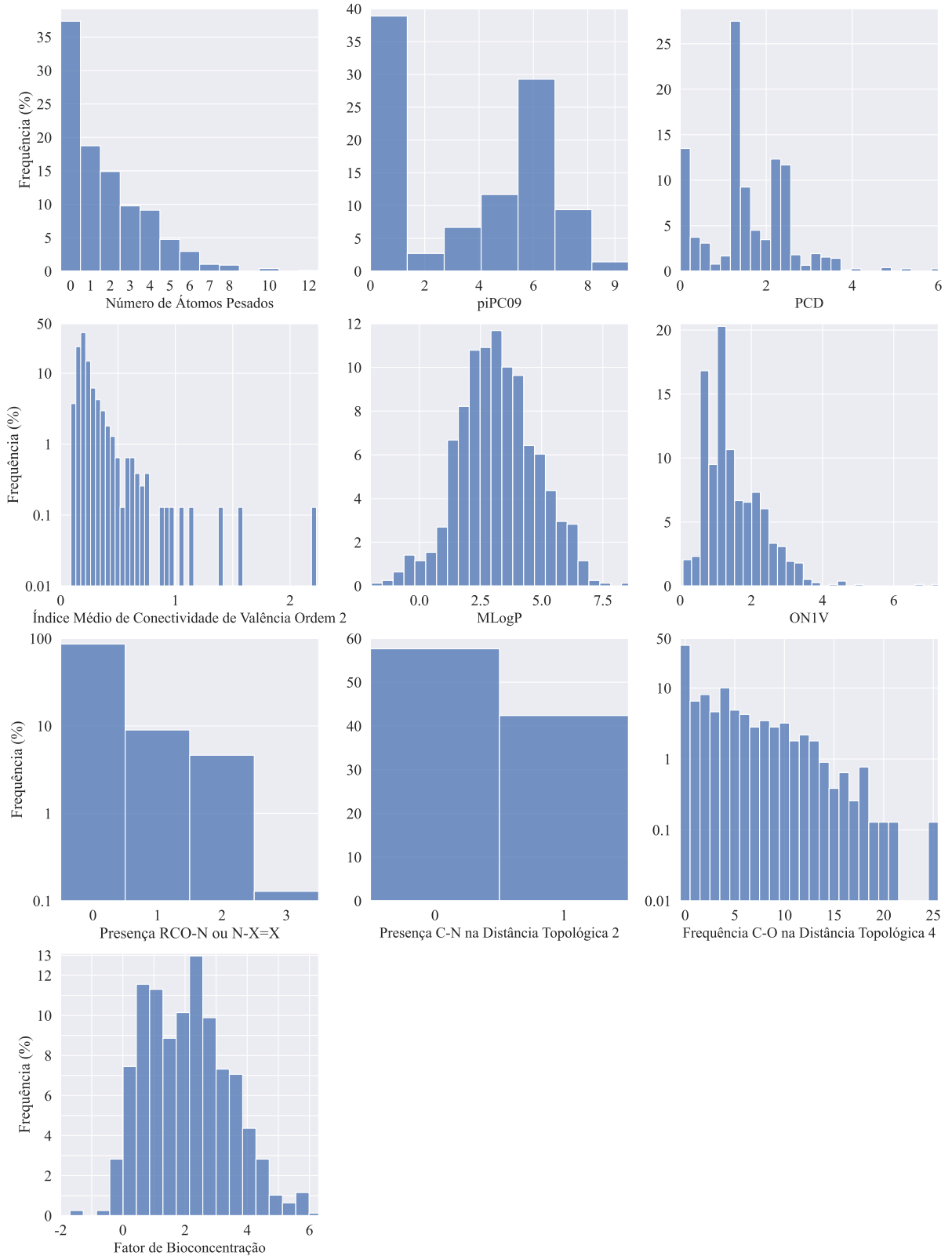
Modelo	Escória Alto-Forno (kg/m ³)	Água (kg/m ³)	Cimento (kg/m ³)	Superplasticante (kg/m ³)	Cinzas Volantes (kg/m ³)	Idade (dias)	Agregado Miúdo (kg/m ³)	Agregado Graúdo (kg/m ³)
LASSO	0,404	-0,135	-0,123	-0,054	-0,159	-0,075	-0,042	-0,009
Ridge	0,369	-0,130	-0,201	-0,058	-0,141	-0,062	-0,039	-0,000
Elastic Net	0,339	-0,176	-0,160	-0,085	-0,109	-0,100	0,031	-0,001
Relaxed Lasso	0,396	-0,191	-0,133	-0,057	-0,149	-0,075	0,000	0,000
KNN	-0,138	-0,195	-0,069	-0,220	-0,141	-0,104	-0,047	-0,086
SVM	0,164	-0,179	0,167	-0,139	-0,239	0,079	-0,021	-0,012
Decision Tree	0,411	-0,319	-0,200	0,000	0,000	0,070	0,000	0,000
Random Forest	0,257	-0,264	-0,177	-0,082	-0,001	0,124	0,076	0,020
Extra Trees	0,175	-0,126	-0,103	-0,238	0,008	0,174	0,164	0,011
XGBoost	0,156	-0,300	-0,111	-0,098	-0,001	0,132	0,144	-0,058
LightGBM	0,192	-0,423	-0,027	0,032	-0,017	0,154	0,144	-0,012
CatBoost	0,324	-0,271	-0,167	-0,040	-0,019	0,111	0,030	0,038
MLP	0,073	0,214	-0,162	0,077	-0,229	0,162	-0,014	-0,069
ResNet	-0,271	0,362	-0,137	-0,052	0,025	-0,078	-0,073	0,003
TabNet	0,181	0,053	0,041	0,322	0,213	0,022	0,056	0,112
SAINT	0,121	-0,099	-0,267	-0,229	-0,160	0,009	-0,032	0,084
FastFormer	0,175	0,022	-0,091	0,068	0,035	-0,044	0,336	0,230
FT-Transformer	-0,093	0,176	-0,381	0,076	0,165	-0,091	0,014	-0,003
TabTransformer	-0,028	-0,452	0,055	-0,213	0,074	0,037	-0,128	0,012
Média	0,169	-0,128	-0,118	-0,052	-0,044	0,023	0,031	0,014
Desvio Padrão	0,188	0,207	0,118	0,132	0,121	0,097	0,103	0,068

Tabela 58 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Concreto

Modelo	Idade (dias)	Cimento (kg/m ³)	Escória Alto-Forno (kg/m ³)	Água (kg/m ³)	Superplasticante (kg/m ³)	Cinzas Volantes (kg/m ³)	Agregado Miúdo (kg/m ³)	Agregado Graúdo (kg/m ³)
LASSO	0,175	0,293	0,213	-0,088	0,041	0,133	0,029	0,028
Ridge	0,175	0,292	0,213	-0,087	0,044	0,132	0,029	0,029
Elastic Net	0,202	0,281	0,187	-0,143	0,061	0,096	-0,023	-0,009
Relaxed Lasso	0,199	0,300	0,207	-0,130	0,045	0,119	0,000	0,000
Árvore de Decisão	0,372	0,229	0,059	0,172	0,051	0,067	0,043	0,007
Random Forest	0,350	0,312	0,075	0,107	0,075	0,023	0,038	0,021
Extra Trees	0,372	0,261	0,077	0,079	0,102	0,046	0,041	0,023
XGBoost	0,305	0,163	0,098	0,113	0,130	0,072	0,065	0,054
LightGBM	0,104	0,122	0,135	0,123	0,133	0,111	0,135	0,136
CatBoost	0,304	0,261	0,109	0,137	0,090	0,014	0,053	0,031
TabNet	0,476	0,153	0,049	0,126	0,074	0,031	0,056	0,034
Média Absolutos	0,276	0,242	0,129	0,119	0,077	0,077	0,047	0,034
Desvio Padrão dos Absolutos	0,107	0,063	0,062	0,026	0,032	0,042	0,033	0,035

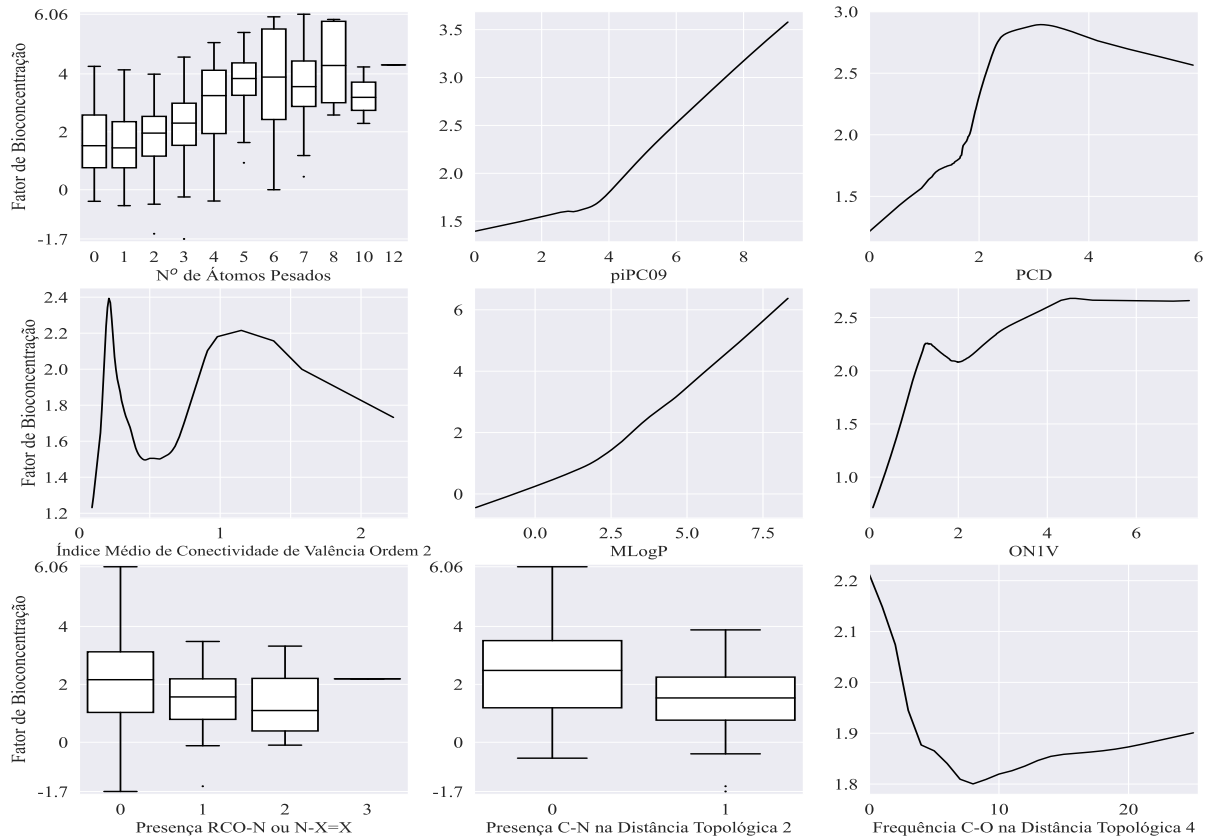
4.12 FATOR DE BIOCONCENTRAÇÃO

Figura 47 – Histogramas das Variáveis para a Base de Dados de Bioconcentração



Fonte: Figura do Autor

Figura 48 – Análise Exploratória da Relação das Variáveis Explicativas com a Resposta para a Base de Dados de Bioconcentração



Fonte: Figura do Autor

Tabela 59 – Métricas para a Base de Dados de Bioconcentração

Modelo	Erro Quadrático Médio	Erro Absoluto Médio	Erro Máximo	R ²	Variância Explicada	Tempo (s)	Razão do Menor Tempo
LASSO	0,626	0,627	3,255	0,644	0,644	0,002	1,8
Ridge	0,618	0,618	3,285	0,649	0,649	0,006	4,6
Elastic Net	0,636	0,635	3,184	0,639	0,639	0,001	1
Relaxed Lasso	0,626	0,627	3,252	0,644	0,644	0,002	1,7
KNN	0,473	0,548	1,886	0,731	0,732	0,002	1,4
SVM	0,434	0,512	2,056	0,753	0,753	0,048	35,5
Decision Tree	0,591	0,595	2,372	0,664	0,665	0,002	1,7
Random Forest	0,523	0,548	2,886	0,703	0,703	1,368	1007,0
Extra Trees	0,482	0,532	2,297	0,726	0,727	0,993	730,4
XGBoost	0,519	0,564	2,461	0,705	0,707	0,064	47,3
LightGBM	0,483	0,533	2,090	0,725	0,727	0,396	291,7
CatBoost	0,453	0,518	1,964	0,743	0,743	4,431	3260,4
MLP	0,535	0,582	2,129	0,696	0,721	1,455	1071,0
ResNet	0,609	0,596	2,877	0,654	0,655	16,833	12387,1
TabNet	0,854	0,747	2,372	0,515	0,517	9,129	6718,0
SAINT	0,578	0,575	2,606	0,672	0,672	6,097	4486,8
FastFormer	0,600	0,603	2,170	0,659	0,659	7,997	5884,5
FT-Transformer	0,689	0,679	2,905	0,608	0,636	4,734	3483,6
Trivial	1,759	1,094	3,773	0	0	-	-

A tabela 59 demonstra o bom desempenho do SVM sobre os outros modelos, situação que não se repetiu nos maiores bancos de dados em estudo, ilustrando a capacidade desse algoritmo de aprender com poucos dados, mas que não há ganhos massivos de desempenho se mais amostras estiverem disponíveis.

Tabela 60 – SHAP Médio para a Base de Dados de Bioconcentração

Modelo	MLogP	Nº de Átomos Pesados	Frequência C-O na Distância Topológica 4	Índice Médio de Conectividade de Valência Ordem 2	piPC09	PCD	Presença RCO-N ou N-X=X	Presença C-N na Distância Topológica 2	ON1V
LASSO	0,629	0,043	0,107	0	-0,079	0,026	-0,071	0,045	0
Ridge	0,419	-0,045	0,143	0,059	0,068	-0,053	0,028	-0,028	0,159
Elastic Net	0,357	0,144	0,132	0	-0,215	0,038	-0,105	0,010	0
Relaxed Lasso	0,284	-0,095	-0,012	0	0,193	-0,187	0,201	-0,028	0
KNN	0,102	0,378	0,237	-0,061	0,035	0,082	-0,002	-0,006	0,097
SVM	-0,240	0,162	0,099	-0,139	-0,050	-0,152	-0,080	0,057	-0,022
Decision Tree	-0,388	0,096	0,016	0,011	-0,245	-0,237	0	0	-0,008
Random Forest	-0,045	0,205	0,019	-0,381	-0,052	-0,148	-0,004	-0,039	0,107
Extra Trees	-0,224	0,207	0,190	-0,155	-0,083	-0,060	0,041	0,022	0,019
XGBoost	-0,234	0,264	0,116	-0,108	0,043	-0,112	-0,009	-0,038	0,076
LightGBM	-0,370	0,192	0,132	0,004	0,005	0,076	0,112	-0,092	-0,018
CatBoost	-0,300	0,069	0,222	-0,301	-0,047	0,002	-0,009	-0,033	-0,017
MLP	-0,017	0,076	0,130	0,066	0,233	-0,133	0,107	0,179	-0,059
ResNet	-0,088	-0,101	0,311	0,017	-0,037	-0,033	0,315	-0,010	-0,089
TabNet	0,043	0,312	0,169	0,336	0,034	0,018	0,055	0,020	-0,014
SAINT	-0,354	-0,221	-0,055	0,002	-0,119	-0,101	-0,018	-0,094	-0,035
FastFormer	-0,328	-0,123	-0,060	0,095	-0,114	-0,032	-0,145	0,079	-0,024
FT-Transformer	-0,042	-0,226	0,254	0,014	-0,016	-0,018	0,189	0,153	-0,088
TabTransformer	0,770	-0,002	-0,009	0,007	-0,071	-0,053	0,003	0,007	0,077
Média	-0,001	0,070	0,113	-0,028	-0,027	-0,057	0,032	0,011	0,008
Desvio Padrão	0,338	0,169	0,104	0,147	0,114	0,086	0,110	0,068	0,064

Tabela 61 – SHAP Absolutos para a Base de Dados de Bioconcentração

Modelo	MLogP	Nº de Átomos Pesados	piPC09	Frequência C-O na Distância Topológica 4	Índice Médio de Conectividade de Valência Ordem 2	PCD	Presença C-N na Distância Topológica 2	ON1V	Presença RCO-N ou N-X=X
LASSO	0,659	0,151	0,099	0,023	0	0,008	0,045	0	0,014
Ridge	0,522	0,168	0,077	0,054	0,005	0,041	0,063	0,047	0,024
Elastic Net	0,649	0,159	0,099	0,011	0	0,017	0,057	0	0,009
Relaxed Lasso	0,658	0,151	0,099	0,023	0	0,008	0,046	0	0,014
KNN	0,282	0,246	0,095	0,111	0,047	0,083	0,052	0,076	0,008
SVM	0,458	0,189	0,049	0,037	0,063	0,082	0,057	0,042	0,023
Decision Tree	0,610	0,057	0,146	0,010	0,006	0,141	0	0,030	0
Random Forest	0,656	0,081	0,059	0,033	0,062	0,055	0,007	0,044	0,002
Extra Trees	0,479	0,138	0,132	0,044	0,044	0,042	0,074	0,023	0,023
XGBoost	0,617	0,085	0,065	0,041	0,062	0,075	0,012	0,039	0,003
LightGBM	0,472	0,111	0,053	0,072	0,067	0,073	0,050	0,059	0,043
CatBoost	0,461	0,124	0,070	0,084	0,114	0,063	0,020	0,060	0,004
MLP	0,416	0,146	0,066	0,057	0,048	0,049	0,064	0,060	0,093
ResNet	0,287	0,135	0,037	0,166	0,072	0,039	0,068	0,091	0,106
TabNet	0,043	0,312	0,034	0,169	0,336	0,018	0,020	0,014	0,055
SAINT	0,397	0,115	0,110	0,077	0,075	0,074	0,073	0,052	0,026
FastFormer	0,438	0,072	0,065	0,078	0,083	0,028	0,103	0,040	0,094
FT-Transformer	0,333	0,193	0,033	0,115	0,073	0,034	0,060	0,075	0,083
TabTransformer	0,785	0,005	0,110	0,006	0,005	0,020	0,005	0,059	0,005
Média	0,485	0,139	0,079	0,064	0,061	0,050	0,046	0,043	0,033
Desvio Padrão	0,171	0,067	0,032	0,047	0,073	0,032	0,027	0,026	0,034

Tabela 62 – Feature Permutation para a Base de Dados de Bioconcentração

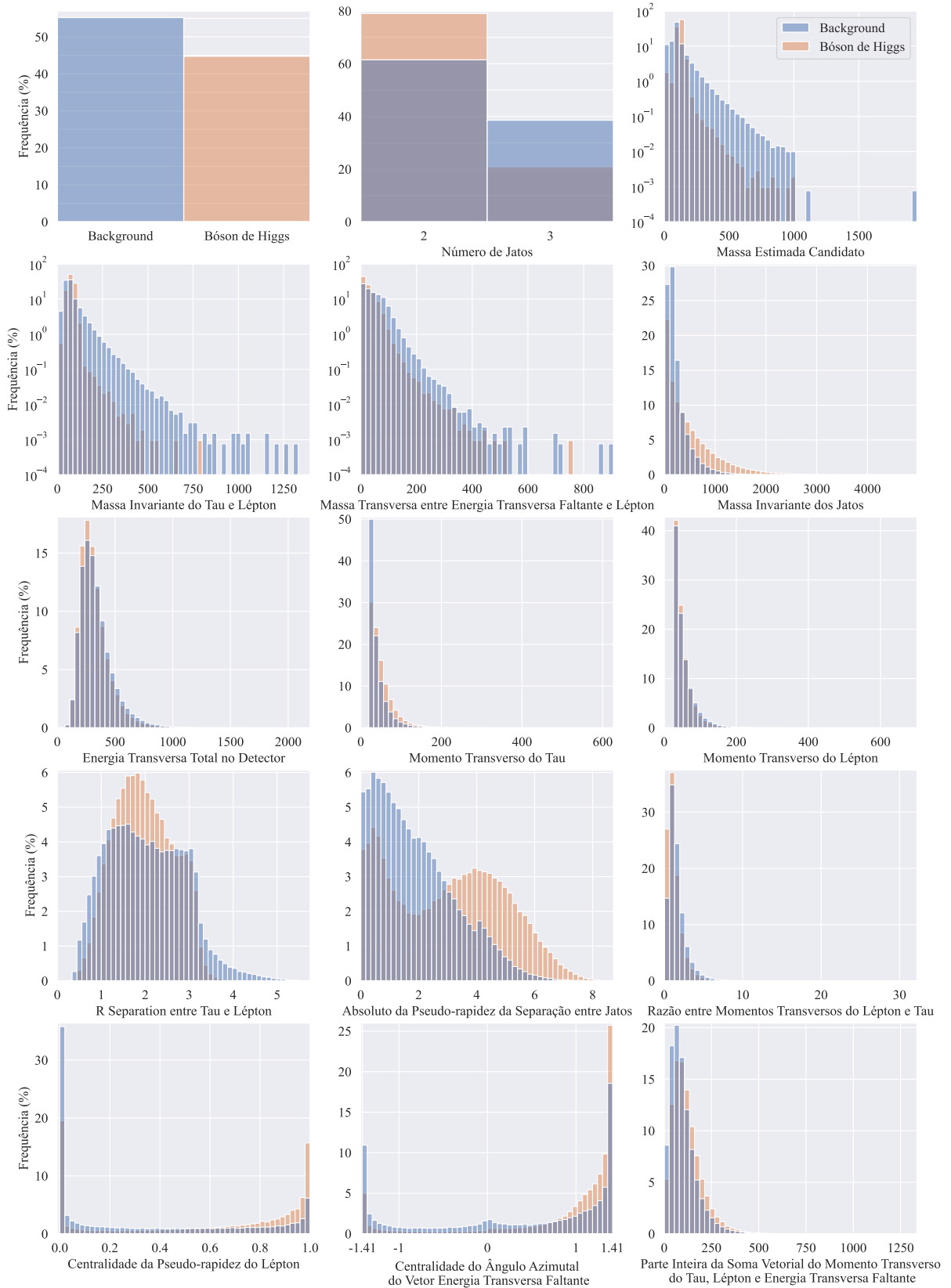
Modelo	MLogP	ONIV	piPC09	Nº de Átomos Pesados	Presença C-N na Distância Topológica 2	PCD	Frequência C-O na Distância Topológica 4	Índice Médio de Conectividade de Valência Ordem 2	Presença RCO-N ou N-X=X
LASSO	0,605	0	0,127	-0,209	0,043	-0,003	0,007	0	0,006
Ridge	0,514	0,124	0,105	-0,173	0,054	-0,004	0,012	0,001	0,014
Elastic Net	0,581	0	0,124	-0,226	0,053	-0,007	0,002	0	0,007
Relaxed Lasso	0,655	0	0,119	-0,161	0,042	-0,003	0,014	0	0,007
KNN	0,390	0,149	-0,094	0,000	0,010	0,066	-0,033	0,145	0,113
SVM	0,340	-0,196	-0,272	0,023	0,066	-0,021	-0,053	0,001	0,029
Decision Tree	0,513	-0,201	0	0,285	0	0	0	0	0
Random Forest	0,371	-0,346	-0,102	0,014	0,006	-0,078	-0,033	-0,021	0,028
Extra Trees	0,577	-0,076	-0,144	-0,119	-0,012	-0,021	-0,028	-0,012	0,011
XGBoost	0,436	-0,107	-0,176	-0,092	0,032	-0,081	-0,029	-0,025	0,022
LightGBM	0,324	-0,126	-0,191	-0,104	-0,042	-0,051	-0,116	-0,035	0,011
CatBoost	0,332	-0,152	-0,269	-0,001	0,040	-0,079	-0,082	-0,013	0,031
MLP	0,406	0,092	-0,219	-0,008	0,093	0,073	0,010	-0,014	0,085
ResNet	-0,057	-0,062	-0,137	0,134	0,144	-0,086	0,098	0,253	0,030
TabNet	0,111	-0,016	0,265	0,090	-0,139	0,152	0,107	-0,068	-0,051
SAINT	0,173	-0,163	-0,108	0,059	0,127	-0,172	0,081	-0,110	0,006
FastFormer	0,400	-0,039	-0,044	0,126	0,248	0,025	0,028	0,087	0,003
FT-Transformer	0,076	0,247	-0,051	0,084	0,047	-0,185	-0,169	-0,081	-0,060
TabTransformer	-0,169	-0,756	0,016	-0,001	0,000	-0,058	-0,000	0,000	-0,000
Média	0,346	-0,086	-0,055	-0,015	0,043	-0,028	-0,010	0,006	0,015
Desvio Padrão	0,221	0,208	0,146	0,129	0,077	0,078	0,066	0,079	0,037

Tabela 63 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados de Bioconcentração

Modelo	MLogP	Nº de Átomos Pesados	piPC09	Presença C-N na Distância Topológica 2	Índice Médio de Conectividade de Valência Ordem 2	Frequência C-O na Distância Topológica 4	ONIV	PCD	Presença RCO-N ou N-X=X
LASSO	0,681	0,145	0,082	-0,036	0,000	-0,025	0,000	0,009	-0,021
Ridge	0,532	0,159	0,063	-0,049	-0,007	-0,058	0,051	0,044	-0,035
Elastic Net	0,674	0,153	0,083	-0,046	-0,000	-0,012	0,000	0,018	-0,014
Relaxed Lasso	0,680	0,145	0,082	-0,036	0,000	-0,025	0,000	0,009	-0,021
Decision Tree	0,897	0,058	0,008	0,000	0,001	0,002	0,023	0,011	0,000
Random Forest	0,732	0,046	0,062	0,003	0,030	0,025	0,060	0,035	0,006
Extra Trees	0,485	0,116	0,135	0,078	0,025	0,042	0,042	0,063	0,014
XGBoost	0,278	0,155	0,112	0,212	0,040	0,039	0,066	0,053	0,044
LightGBM	0,133	0,075	0,123	0,158	0,079	0,121	0,122	0,113	0,076
CatBoost	0,379	0,105	0,087	0,060	0,068	0,085	0,104	0,091	0,022
TabNet	0,159	0,211	0,012	0,065	0,410	0,118	0,016	0,000	0,008
Média Absolutos	0,512	0,124	0,077	0,068	0,060	0,050	0,044	0,041	0,024
Desvio Padrão dos Absolutos	0,238	0,047	0,038	0,061	0,114	0,039	0,040	0,035	0,021

4.13 DETECÇÃO DO BÓSON DE HIGGS

Figura 49 – Histogramas das Variáveis para a Base de Dados do Bóson de Higgs



Fonte: Figura do Autor

Tabela 64 – Métricas para a Base de Dados do Bóson de Higgs

Modelo	Acurácia Balançada	F0,5	F1	F2	AUROC	Recall	Precisão	AUPR	Log Loss	Brier Score	Kappa Cohen	MCC	Tempo (s)	Razão do Menor Tempo
LASSO	0,729	0,731	0,731	0,731	0,809	0,731	0,731	0,776	0,528	0,177	0,457	0,457	162,8	87,9
Ridge	0,729	0,731	0,731	0,731	0,809	0,731	0,731	0,776	0,528	0,177	0,457	0,457	6,2	3,4
Elastic Net	0,729	0,731	0,731	0,731	0,809	0,731	0,731	0,776	0,528	0,177	0,457	0,457	168,4	90,9
KNN	0,812	0,812	0,811	0,811	0,891	0,811	0,813	0,874	0,450	0,136	0,620	0,621	1,9	1
SVM Linear	0,729	0,731	0,731	0,731	0,808	0,731	0,731	0,776	0,529	0,177	0,457	0,457	26,9	14,5
Decision Tree	0,818	0,819	0,819	0,818	0,898	0,818	0,819	0,881	0,400	0,127	0,634	0,635	5,4	2,9
Random Forest	0,839	0,840	0,840	0,840	0,917	0,840	0,840	0,906	0,374	0,117	0,676	0,676	522,1	281,9
Extra Trees	0,839	0,839	0,839	0,839	0,916	0,839	0,840	0,905	0,379	0,118	0,675	0,675	351,7	189,9
XGBoost	0,852	0,853	0,853	0,853	0,929	0,853	0,853	0,920	0,337	0,105	0,703	0,703	22,8	12,3
LightGBM	0,855	0,856	0,855	0,855	0,931	0,855	0,856	0,922	0,331	0,103	0,708	0,708	21,2	11,4
CatBoost	0,855	0,856	0,856	0,855	0,931	0,855	0,856	0,922	0,332	0,104	0,708	0,708	91,7	49,5
MLP	0,846	0,846	0,845	0,845	0,923	0,845	0,847	0,913	0,350	0,110	0,689	0,690	256,3	138,4
ResNet	0,854	0,855	0,854	0,854	0,931	0,854	0,855	0,921	0,332	0,104	0,706	0,706	125,1	67,6
TabNet	0,855	0,856	0,856	0,856	0,932	0,856	0,856	0,922	0,333	0,104	0,709	0,709	334,4	180,6
SAINT	0,853	0,853	0,853	0,852	0,930	0,852	0,854	0,921	0,337	0,105	0,703	0,704	238,7	128,9
FastFormer	0,855	0,855	0,855	0,855	0,931	0,854	0,855	0,922	0,333	0,104	0,707	0,707	374,5	202,2
FT-Transformer	0,855	0,856	0,856	0,856	0,930	0,856	0,856	0,921	0,335	0,105	0,708	0,708	350,5	189,2
TabTransformer	0,854	0,854	0,853	0,853	0,930	0,853	0,854	0,921	0,334	0,104	0,704	0,705	167,7	90,5
Trivial	0,5	0,335	0,393	0,475	0,5	0,552	0,305	0,724	0,688	0,247	0	0	-	-

É possível notar na tabela 64 a superioridade do TabNet na maior parte das métricas, fato que somente ocorreu no banco de dados de maior quantidade de amostras, indicando a capacidade desses algoritmos baseados em redes neurais de aprenderem com maiores volumes de dados.

Tabela 65 – SHAP Médio para a Base de Dados do Bóson de Higgs

Modelo	Massa Estimada Candidato	Massa Invariante do τ e Lépton	Massa Invariante dos Jatos	Absoluto da Pseudo-rapidez da Separação entre Jatos	Centralidade do Ângulo Azimutal do Vetor Energia Transversa Faltante	Momento Transverso do τ	R Separation entre τ e Lépton	Massa Transversa entre Energia Transversa Faltante e Lépton
LASSO	0,000	0,097	-0,225	0,029	0,025	-0,014	-0,001	0,096
Ridge	-0,000	0,094	-0,219	0,028	0,024	-0,013	-0,002	0,094
Elastic Net	-0,000	0,098	-0,225	0,029	0,025	-0,014	-0,002	0,096
KNN	-0,139	-0,124	0,025	0,049	-0,024	-0,037	-0,101	-0,046
SVM Linear	0,001	0,077	-0,250	0,032	0,024	-0,015	-0,001	0,089
Decision Tree	-0,514	-0,104	0,005	0,067	-0,089	-0,094	-0,056	-0,014
Random Forest	-0,451	-0,083	0,030	0,060	0,050	-0,076	-0,022	-0,054
Extra Trees	-0,461	-0,146	0,034	0,080	0,045	-0,060	-0,022	-0,016
XGBoost	-0,326	-0,096	0,042	0,040	0,072	-0,074	0,004	-0,037
LightGBM	-0,310	-0,089	0,061	0,050	0,071	-0,040	0,022	-0,005
CatBoost	-0,339	-0,088	0,079	0,031	0,060	-0,043	0,033	-0,002
MLP	-0,226	-0,301	0,018	0,084	0,019	-0,037	-0,074	-0,031
ResNet	-0,199	-0,229	0,045	0,056	0,051	-0,040	-0,088	-0,008
TabNet	0,333	0,084	0,062	0,098	-0,001	-0,001	0,110	0,061
SAINT	-0,247	-0,182	0,070	0,097	0,061	-0,045	-0,039	-0,022
FastFormer	-0,264	-0,142	0,075	0,071	0,061	-0,065	-0,047	-0,017
FT-Transformer	-0,298	-0,158	0,103	0,090	0,048	-0,049	-0,041	-0,001
TabTransformer	-0,210	-0,244	0,036	0,071	0,057	-0,038	-0,058	-0,008
Média	-0,203	-0,085	-0,013	0,059	0,032	-0,042	-0,021	0,010
Desvio Padrão	0,203	0,122	0,118	0,024	0,038	0,024	0,048	0,051

Tabela 66 – SHAP Absolutos para a Base de Dados do Bóson de Higgs

Modelo	Massa Estimada Candidato	Massa Invariante do τ e Lépton	R Separation entre τ e Lépton	Centralidade do Ângulo Azimutal do Vetor Energia Transversa Faltante	Centralidade da Pseudo-rapidez do Lépton	Massa Invariante dos Jatos	Parte Inteira da Soma Vetorial do Momento Transverso do τ , Lépton e Energia Transversa Faltante	Momento Transverso do τ
LASSO	0,019	0,070	0,104	0,041	0,063	0,125	0,084	0,051
Ridge	0,019	0,069	0,103	0,040	0,062	0,123	0,082	0,048
Elastic Net	0,019	0,070	0,104	0,041	0,063	0,125	0,084	0,051
KNN	0,090	0,071	0,059	0,087	0,051	0,024	0,042	0,054
SVM Linear	0,016	0,063	0,097	0,044	0,067	0,129	0,067	0,054
Decision Tree	0,449	0,085	0,097	0,082	0,028	0,006	0,004	0,085
Random Forest	0,361	0,070	0,043	0,091	0,063	0,029	0,032	0,068
Extra Trees	0,280	0,087	0,035	0,135	0,100	0,026	0,017	0,053
XGBoost	0,262	0,073	0,032	0,078	0,077	0,044	0,037	0,061
LightGBM	0,255	0,058	0,046	0,067	0,067	0,041	0,028	0,043
CatBoost	0,258	0,058	0,044	0,059	0,078	0,049	0,025	0,045
MLP	0,144	0,131	0,061	0,045	0,066	0,044	0,099	0,058
ResNet	0,130	0,093	0,066	0,056	0,028	0,040	0,061	0,048
TabNet	0,333	0,084	0,110	0,001	0,017	0,062	0,002	0,001
SAINT	0,157	0,104	0,049	0,068	0,057	0,037	0,049	0,049
FastFormer	0,199	0,076	0,048	0,061	0,072	0,042	0,063	0,061
FT-Transformer	0,174	0,071	0,038	0,054	0,068	0,046	0,061	0,043
TabTransformer	0,133	0,097	0,048	0,049	0,048	0,038	0,084	0,043
Média	0,183	0,079	0,066	0,061	0,060	0,057	0,051	0,051
Desvio Padrão	0,124	0,018	0,027	0,027	0,019	0,038	0,028	0,016

Tabela 67 – Feature Permutation para a Base de Dados do Bóson de Higgs

Modelo	Massa Transversa entre Energia Transversa Faltante e Lépton	Centralidade do Ângulo Azimutal do Vetor Energia Transversa Faltante	Centralidade da Pseudo-rapidez do Lépton	Absoluto da Pseudo-rapidez da Separação entre Jatos	Massa Invariante do τ e Lépton	Massa Estimada Candidato	Razão entre Momentos Transversos do Lépton e τ	Momento Transverso do Lépton
LASSO	-0,186	-0,120	0,043	-0,040	-0,100	-0,015	-0,172	0,029
Ridge	-0,170	-0,095	0,038	-0,030	-0,136	0,003	-0,184	0,089
Elastic Net	-0,181	-0,104	0,061	-0,055	-0,158	0,011	-0,158	0,087
KNN	-0,229	-0,174	0,023	-0,013	-0,009	0,005	-0,098	-0,063
SVM Linear	-0,187	-0,117	0,054	-0,024	-0,117	0,001	-0,173	0,100
Decision Tree	-0,184	-0,220	0,048	0,364	0	0,132	0	0
Random Forest	-0,223	-0,112	0,118	0,155	-0,012	0,168	-0,011	-0,001
Extra Trees	-0,288	-0,272	0,149	0,040	0,002	0,070	-0,026	-0,002
XGBoost	-0,237	-0,136	0,069	0,052	-0,017	0,092	-0,003	0,067
LightGBM	-0,282	-0,148	0,053	0,036	-0,093	0,037	-0,025	0,044
CatBoost	-0,274	-0,123	0,030	0,024	-0,165	0,028	-0,005	0,015
MLP	-0,212	-0,037	0,051	0,027	-0,148	0,027	-0,151	0,079
ResNet	-0,018	0,038	0,059	0,063	0,092	0,020	0,014	0,105
TabNet	0,040	0,071	-0,017	0,029	-0,081	-0,223	0,004	-0,090
SAINT	-0,066	-0,005	0,155	0,125	-0,013	-0,103	-0,002	0,027
FastFormer	-0,059	0,010	0,194	0,142	0,022	0,070	-0,013	0,035
FT-Transformer	-0,035	0,004	0,173	0,114	0,068	0,093	-0,007	0,014
TabTransformer	-0,025	0,023	0,119	0,105	0,025	0,016	0,001	0,074
Média	-0,156	-0,084	0,079	0,062	-0,047	0,024	-0,056	0,034
Desvio Padrão	0,099	0,091	0,057	0,096	0,077	0,084	0,073	0,052

Tabela 68 – Importância das Variáveis Internamente a cada Modelo para a Base de Dados do Bóson de Higgs

Modelo	Massa Estimada Candidato	Massa Invariante do τ e Lépton	Massa Invariante dos Jatos	Absoluto da Pseudo-rapidez da Separação entre Jatos	Centralidade do Ângulo Azimutal do Vetor Energia Transversa Faltante	R Separation entre τ e Lépton	Centralidade da Pseudo-rapidez do Lépton	Massa Transversa entre Energia Transversa Faltante e Lépton
LASSO	0,026	-0,091	0,134	-0,044	0,035	0,093	0,049	-0,052
Ridge	0,025	-0,090	0,132	-0,043	0,034	0,091	0,048	-0,051
Elastic Net	0,026	-0,091	0,135	-0,044	0,035	0,093	0,049	-0,052
Decision Tree	0,441	0,112	0,009	0,185	0,081	0,037	0,039	0,033
Random Forest	0,416	0,101	0,051	0,080	0,050	0,029	0,053	0,054
Extra Trees	0,302	0,089	0,046	0,097	0,093	0,039	0,116	0,044
XGBoost	0,173	0,085	0,052	0,063	0,067	0,040	0,073	0,042
LightGBM	0,039	0,035	0,030	0,027	0,044	0,034	0,040	0,047
CatBoost	0,146	0,062	0,048	0,037	0,045	0,052	0,039	0,056
TabNet	0,173	0,095	0,044	0,062	0,089	0,041	0,044	0,061
Média Absolutos	0,177	0,085	0,068	0,068	0,057	0,055	0,055	0,049
Desvio Padrão dos Absolutos	0,152	0,021	0,045	0,044	0,022	0,025	0,022	0,008

5 RESULTADOS TESTES DE HIPÓTESE

Para o banco de dados dos Atrasos no SUS o Teste Q de Cochran obteve uma estatística com um valor acima de 2675, com seu respectivo p-valor menor do que 10^{-15} , levando a hipótese nula, definida como não haver diferença estatisticamente significativa entre a acurácia dos algoritmos, a ser rejeitada para um nível de significância de 5%. Então foram realizados testes de McNemar *post-hoc* para cada par das redes neurais.

Tabela 69 – Testes de Hipótese para a Base de Dados dos Atrasos no SUS

Rede Neural	Rede Neural	Qui-Quadrado	P-Valor	Holm-Bonferroni	Hipótese Nula
MLP	ResNet	9612,0001	0,0000	0,0024	Rejeitada
MLP	TabNet	8925,0001	0,0000	0,0025	Rejeitada
MLP	SAINT	9536,0001	0,0000	0,0026	Rejeitada
MLP	FastFormer	9961,0001	0,0000	0,0028	Rejeitada
MLP	FT-Transformer	9798,0001	0,0000	0,0029	Rejeitada
MLP	TabTransformer	8430,0001	0,0000	0,0031	Rejeitada
FastFormer	TabTransformer	954,3009	$1,5412 \times 10^{-209}$	0,0033	Rejeitada
FT-Transformer	TabTransformer	836,4767	$6,3326 \times 10^{-184}$	0,0036	Rejeitada
ResNet	TabTransformer	748,2623	$9,5771 \times 10^{-165}$	0,0038	Rejeitada
SAINT	TabTransformer	615,4360	$7,3522 \times 10^{-136}$	0,0042	Rejeitada
TabNet	FastFormer	487,8074	$4,2747 \times 10^{-108}$	0,0045	Rejeitada
TabNet	FT-Transformer	405,5381	$3,4306 \times 10^{-90}$	0,0050	Rejeitada
ResNet	TabNet	316,4734	$8,4933 \times 10^{-71}$	0,0056	Rejeitada
TabNet	TabTransformer	244,2803	$4,5861 \times 10^{-55}$	0,0062	Rejeitada
TabNet	SAINT	180,7188	$3,3766 \times 10^{-41}$	0,0071	Rejeitada
ResNet	FastFormer	109,3984	$1,3274 \times 10^{-25}$	0,0083	Rejeitada
SAINT	FastFormer	92,8116	$5,7518 \times 10^{-22}$	0,0100	Rejeitada
SAINT	FT-Transformer	35,8909	$2,0868 \times 10^{-9}$	0,0125	Rejeitada
ResNet	FT-Transformer	34,7815	$3,6886 \times 10^{-9}$	0,0167	Rejeitada
FastFormer	FT-Transformer	22,9606	$1,6535 \times 10^{-6}$	0,0250	Rejeitada
ResNet	SAINT	3,0112	$8,2689 \times 10^{-2}$	0,0500	Não Rejeitada

Sendo representado, nas tabelas 70 e 72, em azul os testes de hipótese, entre cada par, que rejeitam a hipótese nula e em vermelho os que não rejeitam.

Tabela 70 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados dos Atrasos no SUS

	MLP	ResNet	TabNet	TabTransformer	FT-Transformer	FastFormer	SAINT
MLP	Blue	Blue	Blue	Blue	Blue	Blue	Blue
ResNet	Blue	White	Blue	Blue	Blue	Blue	Red
TabNet	Blue	Blue	White	Blue	Blue	Blue	Blue
TabTransformer	Blue	Blue	Blue	White	Blue	Blue	Blue
FT-Transformer	Blue	Blue	Blue	Blue	White	Blue	Blue
FastFormer	Blue	Blue	Blue	Blue	Blue	White	Blue
SAINT	Blue	Red	Blue	Blue	Blue	Blue	White

Para o banco de dados das Pulsares o teste Q de Cochran retornou uma estatística com um valor superior a 193, com respectivo p-valor inferior a 10^{-15} , rejeitando a hipótese nula para um nível de significância de 5%.

Tabela 71 – Testes de Hipótese para a Base de Dados das Pulsares

Rede Neural	Rede Neural	Qui-Quadrado	P-Valor	Holm-Bonferroni	Hipótese Nula
MLP	ResNet	2598,0004	0,0000	0,0024	Rejeitada
MLP	TabNet	2581,0004	0,0000	0,0025	Rejeitada
MLP	SAINT	2595,0004	0,0000	0,0026	Rejeitada
MLP	FastFormer	2534,0004	0,0000	0,0028	Rejeitada
MLP	FT-Transformer	2606,0004	0,0000	0,0029	Rejeitada
MLP	TabTransformer	2615,0004	0,0000	0,0031	Rejeitada
FastFormer	TabTransformer	58,7156	$1,8219 \times 10^{-14}$	0,0033	Rejeitada
FastFormer	FT-Transformer	58,6163	$1,9162 \times 10^{-14}$	0,0036	Rejeitada
ResNet	FastFormer	40,5000	$1,9662 \times 10^{-10}$	0,0038	Rejeitada
SAINT	FastFormer	37,1134	$1,1146 \times 10^{-9}$	0,0042	Rejeitada
TabNet	FastFormer	24,8941	$6,0567 \times 10^{-7}$	0,0045	Rejeitada
TabNet	TabTransformer	20,9423	$4,7332 \times 10^{-6}$	0,0050	Rejeitada
TabNet	FT-Transformer	15,5676	$7,9608 \times 10^{-5}$	0,0056	Rejeitada
SAINT	TabTransformer	9,5000	$2,0547 \times 10^{-3}$	0,0062	Rejeitada
ResNet	TabTransformer	9,4815	$2,0756 \times 10^{-3}$	0,0071	Rejeitada
ResNet	TabNet	6,2439	$1,2462 \times 10^{-2}$	0,0083	Não Rejeitada
TabNet	SAINT	4,2250	$3,9833 \times 10^{-2}$	0,0100	Não Rejeitada
SAINT	FT-Transformer	3,0303	$8,1723 \times 10^{-2}$	0,0125	Não Rejeitada
FT-Transformer	TabTransformer	1,6410	$2,0018 \times 10^{-1}$	0,0167	Não Rejeitada
ResNet	FT-Transformer	1,4412	$2,2995 \times 10^{-1}$	0,0250	Não Rejeitada
ResNet	SAINT	0,1026	$7,4877 \times 10^{-1}$	0,0500	Não Rejeitada

Tabela 72 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados das Pulsares

	MLP	ResNet	TabNet	TabTransformer	FT-Transformer	FastFormer	SAINT
MLP							
ResNet							
TabNet							
TabTransformer							
FT-Transformer							
FastFormer							
SAINT							

Para o banco de dados do marketing do banco português o teste Q de Cochran possui uma estatística igual a 10,4644 e p-valor de 0,1064 não sendo possível rejeitar a hipótese nula para um nível de significância de 5% e, portanto, não podemos concluir que há diferença estatisticamente significativa entre a performance das redes neurais nessa situação.

Para o teste Q de Cochran do *dataset* do Bóson de Higgs, sua estatística representa um valor de 23,4879 com respectivo p-valor igual a 0,0006. Portanto, rejeitamos a hipótese nula para um nível de significância de 5%.

Tabela 73 – Testes de Hipótese para a Base de Dados do Bóson de Higgs

Rede Neural	Rede Neural	Qui-Quadrado	P-Valor	Holm-Bonferroni	Hipótese Nula
MLP	ResNet	30483,0000	0,0000	0,0024	Rejeitada
MLP	TabNet	30546,0000	0,0000	0,0025	Rejeitada
MLP	SAINT	30428,0000	0,0000	0,0026	Rejeitada
MLP	FastFormer	30501,0000	0,0000	0,0028	Rejeitada
MLP	FT-Transformer	30541,0000	0,0000	0,0029	Rejeitada
MLP	TabTransformer	30449,0000	0,0000	0,0031	Rejeitada
TabNet	SAINT	9,3377	0,0022	0,0033	Rejeitada
SAINT	FT-Transformer	9,1362	0,0025	0,0036	Rejeitada
TabNet	TabTransformer	7,6481	0,0057	0,0038	Não Rejeitada
FT-Transformer	TabTransformer	6,2076	0,0127	0,0042	Não Rejeitada
SAINT	FastFormer	4,5196	0,0335	0,0045	Não Rejeitada
ResNet	TabNet	3,0484	0,0808	0,0050	Não Rejeitada
ResNet	FT-Transformer	2,2752	0,1315	0,0056	Não Rejeitada
FastFormer	TabTransformer	2,1784	0,1400	0,0062	Não Rejeitada
ResNet	SAINT	2,0463	0,1526	0,0071	Não Rejeitada
TabNet	FastFormer	1,6733	0,1958	0,0083	Não Rejeitada
FastFormer	FT-Transformer	1,2386	0,2657	0,0100	Não Rejeitada
ResNet	TabTransformer	1,0782	0,2991	0,0125	Não Rejeitada
SAINT	TabTransformer	0,3177	0,5730	0,0167	Não Rejeitada
ResNet	FastFormer	0,2240	0,6360	0,0250	Não Rejeitada
TabNet	FT-Transformer	0,0127	0,9103	0,0500	Não Rejeitada

Tabela 74 – Representação dos Resultados dos Testes de Hipótese para a Base de Dados do Bóson de Higgs

	MLP	ResNet	TabNet	TabTransformer	FT-Transformer	FastFormer	SAINT
MLP	Blue	Red	Red	Red	Red	Red	Red
ResNet	Blue	White	Red	Red	Red	Red	Red
TabNet	Blue	Red	White	Red	Red	Red	Blue
TabTransformer	Blue	Red	Red	White	Red	Red	Red
FT-Transformer	Blue	Red	Red	Red	White	Red	Blue
FastFormer	Blue	Red	Red	Red	Red	White	Red
SAINT	Blue	Red	Blue	Red	Blue	Red	White

6 IMPORTÂNCIA DOS HIPERPARÂMETROS

As relevâncias dos hiperparâmetros foram calculadas primeiramente ajustando um modelo *Random Forest* para as redes neurais mais complexas, de cada base de dados, tratando os hiperparâmetros como variáveis explicativas e a duração do treino ou a métrica de avaliação como a variável resposta.

Então foi realizado um processo de fANOVA, como descrito por Hutter, Hoos e Leyton-Brown (2014), em que a variância dessa nova variável resposta é decomposta em componentes aditivos, representando a contribuição marginal de cada hiperparâmetro para o resultado.

Para todas as tabelas a soma das importâncias dadas resulta em 1. Foram explicitados somente os hiperparâmetros considerados mais relevantes e também foi ressaltado o maior valor para cada banco de dados.

Tabela 75 – Hiperparâmetros mais Relevantes para a Métrica no ResNet

Hiperparâmetro	Bioconcentração	Concreto	Cardiotelegrafia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,800	0,615	0,546	0,265	0,135	0,581	0,615	0,338	0,584	0,253	0,146	0,404	0,445	0,441
Taxa Aprendizado	0,041	0,043	0,074	0,309	0,614	0,040	0,148	0,182	0,058	0,131	0,612	0,307	0,036	0,640
Decréscimo Taxa	0,063	0,187	0,148	0,041	0,018	0,101	0,061	0,224	0,070	0,049	0,106	0,062	0,056	0,731
Dropout Embed. Tamanho 2 ^a	0,020	0,035	0,095	0,066	0,039	0,163	0,014	0,072	0,039	0,035	0,009	0,146	0,039	0,791
Camada Dropout Blocos	0,008	0,016	0,044	0,055	0,059	0,041	0,016	0,060	0,111	0,199	0,042	0,006	0,087	0,848
Concatenar 1 ^a Contínuas	0,045	0,042	0,019	0,056	0,041	0,017	0,037	0,033	0,031	0,223	0,016	0,023	0,033	0,895
Tamanho 1 ^a Camada	0,001	0,001	0,002	0,029	0,009	0,001	0,057	0,011	0,003	0,003	0,030	0,002	0,180	0,921
Batch	0,006	0,028	0,010	0,021	0,013	0,023	0,014	0,016	0,051	0,038	0,015	0,028	0,034	0,944
	0,006	0,017	0,039	0,082	0,016	0,011	0,011	0,033	0,026	0,018	0,012	0,003	0,013	0,966

Tabela 76 – Hiperparâmetros mais Relevantes para o Tempo no ResNet

Hiperparâmetro	Bioconcentração	Concreto	Cardiotelegrafia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Decréscimo Taxa	0,349	0,705	0,452	0,139	0,252	0,347	0,615	0,263	0,802	0,321	0,854	0,577	0,417	0,469
Batch	0,002	0,108	0,090	0,541	0,603	0,042	0,120	0,162	0,018	0,080	0,002	0,083	0,245	0,630
Tamanho 2 ^a Camada	0,117	0,077	0,072	0,063	0,032	0,212	0,127	0,041	0,106	0,330	0,030	0,052	0,119	0,736
Dropout Blocos	0,082	0,034	0,026	0,120	0,022	0,295	0,012	0,073	0,018	0,043	0,021	0,036	0,101	0,804
N ^o Camadas ResNet	0,206	0,007	0,137	0,013	0,055	0,006	0,007	0,021	0,001	0,032	0,005	0,056	0,027	0,848
Otimizador	0,016	0,032	0,028	0,015	0,002	0,016	0,036	0,352	0,011	0,014	0,006	0,017	0,018	0,891
Taxa Aprendizado	0,168	0,011	0,061	0,046	0,010	0,012	0,013	0,019	0,023	0,036	0,038	0,086	0,034	0,934
Tamanho 1 ^a Camada	0,010	0,013	0,048	0,015	0,005	0,022	0,042	0,021	0,005	0,104	0,033	0,025	0,016	0,962

É interessante notar nas tabelas 75 e 76 que mesmo sendo dada a opção de cada ResNet utilizar de uma a quatro camadas, os números de neurônios nas camadas iniciais foram mais relevantes, tanto para a métrica quanto para o tempo de treinamento. Resultado coerente com o esperado, dado que se uma rede neural possui poucos neurônios de entrada, a quantidade de representações dos dados que ela pode conceber é menor e, portanto, menos informação será processada pelo resto da rede.

Tabela 77 – Hiperparâmetros mais Relevantes para a Métrica no TabNet

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,115	0,942	0,262	0,853	0,119	0,086	0,003	0,457	0,685	0,594	0,883	0,617	0,627	0,480
Lambda Esparso	0,007	0,015	0,137	0,013	0,449	0,403	0,055	0,017	0,030	0,063	0,005	0,052	0,097	0,584
Momentum	0,415	0,004	0,006	0,004	0,069	0,008	0,324	0,007	0,239	0,110	0,013	0,122	0,007	0,686
Gamma	0,404	0,002	0,010	0,006	0,042	0,038	0,561	0,070	0,003	0,016	0,003	0,014	0,063	0,781
Taxa Aprendizado	0,032	0,010	0,064	0,074	0,213	0,040	0,004	0,284	0,006	0,052	0,003	0,033	0,024	0,845
Decréscimo Taxa	0,003	0,021	0,054	0,008	0,008	0,292	0,006	0,078	0,008	0,020	0,048	0,011	0,011	0,889
Steps da Rede	0,001	0,000	0,195	0,002	0,011	0,032	0,001	0,005	0,006	0,005	0,007	0,001	0,012	0,911
Batch Virtual	0,005	0,001	0,013	0,004	0,025	0,015	0,001	0,016	0,005	0,051	0,003	0,124	0,008	0,931
Dimensão Saídas	0,002	0,003	0,095	0,012	0,034	0,007	0,035	0,013	0,001	0,009	0,001	0,004	0,007	0,949
Dropout	0,002	0,001	0,095	0,002	0,000	0,017	0,004	0,005	0,007	0,025	0,001	0,002	0,055	0,965
Batch Total	0,007	0,000	0,020	0,015	0,005	0,036	0,003	0,017	0,006	0,009	0,004	0,007	0,059	0,980

Tabela 78 – Hiperparâmetros mais Relevantes para o Tempo no TabNet

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Steps da Rede	0,423	0,228	0,474	0,127	0,527	0,256	0,103	0,250	0,015	0,079	0,133	0,157	0,370	0,242
Dimensão Saídas	0,061	0,093	0,202	0,054	0,033	0,241	0,338	0,271	0,007	0,315	0,045	0,032	0,362	0,400
Batch Virtual	0,190	0,220	0,042	0,405	0,075	0,091	0,030	0,014	0,070	0,135	0,041	0,432	0,019	0,535
Decréscimo Taxa	0,016	0,094	0,020	0,040	0,032	0,040	0,078	0,059	0,326	0,112	0,091	0,109	0,081	0,620
Lambda Esparso	0,065	0,062	0,066	0,037	0,056	0,046	0,036	0,068	0,221	0,160	0,078	0,109	0,043	0,700
Taxa Aprendizado	0,020	0,021	0,033	0,047	0,017	0,039	0,086	0,012	0,219	0,065	0,101	0,045	0,011	0,755
Momentum	0,056	0,064	0,022	0,034	0,041	0,071	0,110	0,065	0,026	0,022	0,089	0,016	0,018	0,804
Gamma	0,028	0,032	0,044	0,018	0,073	0,024	0,028	0,077	0,021	0,038	0,118	0,038	0,009	0,846
Blocos GLU Compartilhados	0,057	0,027	0,011	0,126	0,038	0,076	0,026	0,050	0,028	0,018	0,050	0,001	0,018	0,887
Blocos GLU Dependentes	0,025	0,112	0,019	0,048	0,041	0,043	0,026	0,069	0,014	0,003	0,095	0,003	0,011	0,926
Batch Total	0,006	0,010	0,013	0,017	0,010	0,051	0,110	0,019	0,038	0,023	0,021	0,021	0,009	0,953
Otimizador	0,015	0,025	0,046	0,014	0,013	0,012	0,018	0,026	0,005	0,007	0,085	0,027	0,035	0,978

Tabela 79 – Hiperparâmetros mais Relevantes para a Métrica no TabTransformer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,632	0,019	0,304	0,836	0,942	0,148	0,720	0,820	0,283	0,256	0,112	0,688	0,425	0,476
Taxa Aprendizado	0,127	0,694	0,061	0,049	0,002	0,068	0,023	0,014	0,157	0,063	0,063	0,009	0,182	0,592
Dropout MLP	0,058	0,015	0,072	0,019	0,002	0,050	0,091	0,022	0,063	0,191	0,426	0,189	0,133	0,695
Decréscimo Taxa	0,043	0,008	0,208	0,026	0,005	0,035	0,054	0,039	0,046	0,220	0,035	0,006	0,054	0,754
Dropout Atenção	0,023	0,148	0,059	0,010	0,001	0,041	0,009	0,008	0,095	0,120	0,073	0,007	0,018	0,802
Dropout Embed.	0,014	0,064	0,099	0,007	0,006	0,140	0,016	0,005	0,042	0,009	0,041	0,015	0,052	0,841
Dropout FF	0,014	0,003	0,027	0,009	0,006	0,213	0,007	0,019	0,067	0,007	0,016	0,018	0,061	0,877
Ativação MLP	0,030	0,004	0,025	0,010	0,005	0,123	0,007	0,014	0,037	0,024	0,012	0,018	0,014	0,902
Ativação Transformer	0,013	0,009	0,032	0,007	0,007	0,029	0,020	0,024	0,042	0,021	0,083	0,009	0,019	0,926
Batch	0,005	0,013	0,030	0,010	0,019	0,021	0,007	0,007	0,031	0,013	0,017	0,001	0,001	0,940
Número Blocos	0,005	0,007	0,009	0,001	0,001	0,021	0,006	0,014	0,016	0,022	0,025	0,015	0,002	0,951
Número Heads	0,012	0,005	0,005	0,000	0,002	0,017	0,005	0,005	0,033	0,026	0,018	0,001	0,011	0,962

Tabela 80 – Hiperparâmetros mais Relevantes para o Tempo no TabTransformer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Decréscimo Taxa	0,772	0,254	0,682	0,167	0,284	0,130	0,232	0,323	0,605	0,538	0,130	0,518	0,288	0,379
Batch	0,001	0,071	0,007	0,403	0,302	0,089	0,491	0,236	0,130	0,175	0,050	0,042	0,570	0,576
Dimensão Modelo	0,002	0,002	0,078	0,001	0,005	0,541	0,003	0,002	0,000	0,006	-	-	0,001	0,640
Dropout FF	0,018	0,056	0,013	0,014	0,135	0,046	0,029	0,032	0,066	0,014	0,235	0,023	0,013	0,694
Dropout Atenção	0,024	0,049	0,007	0,204	0,019	0,022	0,011	0,059	0,034	0,031	0,165	0,065	0,002	0,747
Número Blocos	0,080	0,015	0,021	0,030	0,083	0,007	0,002	0,005	0,007	0,048	0,157	0,074	0,078	0,794
Dropout Embed.	0,006	0,039	0,024	0,029	0,070	0,050	0,019	0,116	0,061	0,021	0,029	0,050	0,007	0,834
Taxa Aprendizado	0,013	0,315	0,031	0,018	0,006	0,007	0,018	0,015	0,004	0,028	0,019	0,005	0,004	0,871
Ativação Transformer	0,017	0,068	0,026	0,019	0,021	0,017	0,117	0,025	0,011	0,034	0,050	0,031	0,003	0,905
Dropout MLP	0,012	0,034	0,041	0,027	0,036	0,021	0,007	0,019	0,008	0,036	0,060	0,095	0,011	0,936
Otimizador	0,030	0,033	0,013	0,040	0,004	0,042	0,024	0,082	0,014	0,036	0,012	0,038	0,003	0,965
Número Heads	0,002	0,014	0,002	0,006	0,003	0,009	0,010	0,059	0,004	0,004	0,040	0,004	0,003	0,977

Tabela 81 – Hiperparâmetros mais Relevantes para a Métrica no FT-Transformer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,853	0,807	0,406	0,673	0,309	0,684	0,846	0,688	0,151	0,234	0,334	0,567	0,826	0,568
Taxa Aprendizado	0,072	0,033	0,105	0,070	0,103	0,089	0,042	0,074	0,155	0,294	0,283	0,134	0,046	0,683
Decréscimo Taxa	0,019	0,044	0,226	0,085	0,063	0,051	0,017	0,056	0,407	0,030	0,016	0,017	0,017	0,763
Dropout FF	0,005	0,006	0,020	0,004	0,229	0,008	0,007	0,021	0,061	0,144	0,056	0,032	0,006	0,809
Dropout Embed.	0,005	0,003	0,108	0,016	0,009	0,035	0,010	0,017	0,022	0,086	0	0,100	0,009	0,844
Dropout Atenção	0,000	0,034	0,011	0,066	0,039	0,025	0,008	0,024	0,031	0,016	0,089	0,039	0,008	0,874
Compressão kv	0,010	0,014	0,017	0,015	0,125	0,017	0,007	0,014	0,022	0,019	0,033	0,013	0,017	0,899
Fator FF	0,014	0,007	0,025	0,012	0,023	0,015	0,004	0,027	0,009	0,015	0,026	0,023	0,024	0,917
Ativ. Embed. Contínuos	0,014	0,028	0,011	0,019	0,017	0,009	0,010	0,023	0,026	0,010	0,013	0,037	0,007	0,934
Ativação Transformer	0,005	0,005	0,024	0,005	0,023	0,016	0,008	0,020	0,058	0,025	0,009	0,007	0,015	0,951
Número Blocos	0,000	0,001	0,010	0,012	0,005	0,025	0,004	0,008	0,019	0,009	0,063	0,015	0,008	0,964
Número Heads	0,001	0,003	0,005	0,004	0,039	0,005	0,005	0,007	0,009	0,032	0,031	0,001	0,002	0,976
Batch	0,000	0,006	0,014	0,003	0,006	0,009	0,026	0,004	0,006	0,033	0,012	0,004	0,005	0,985

Tabela 82 – Hiperparâmetros mais Relevantes para o Tempo no FT-Transformer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Decréscimo Taxa	0,676	0,553	0,768	0,630	0,211	0,485	0,547	0,736	0,273	0,550	0,103	0,193	0,220	0,457
Número Blocos	0,025	0,016	0,027	0,049	0,111	0,206	0,287	0,005	0,218	0,027	0,009	0,004	0,479	0,570
Taxa Aprendizado	0,024	0,022	0,039	0,004	0,008	0,008	0,006	0,013	0,035	0,131	0,668	0,007	0,002	0,644
Batch	0,002	0,058	0,012	0,127	0,457	0,016	0,034	0,056	0,020	0,008	0,003	0,018	0,004	0,707
Compartilhar Embeddings	0,002	0,001	0,002	0,010	0,003	0,009	0,009	0,001	0,078	0,002	0,005	0,589	0,005	0,762
Dropout Atenção	0,007	0,047	0,023	0,009	0,060	0,119	0,011	0,032	0,056	0,012	0,066	0,033	0,133	0,809
Fator FF	0,092	0,161	0,021	0,032	0,004	0,017	0,010	0,031	0,046	0,021	0,012	0,015	0,031	0,847
Compressão kv	0,013	0,076	0,009	0,005	0,020	0,037	0,011	0,005	0,095	0,039	0,009	0,075	0,012	0,878
Dropout FF	0,027	0,009	0,029	0,004	0,025	0,013	0,013	0,009	0,025	0,100	0,081	0,013	0,020	0,906
Otimizador	0,046	0,009	0,031	0,015	0,012	0,017	0,032	0,044	0,022	0,026	0,009	0,029	0,005	0,929
Dropout Embed.	0,012	0,014	0,010	0,040	0,016	0,011	0,006	0,027	0,022	0,016	0	0,009	0,031	0,947
Ativação Transformer	0,033	0,018	0,007	0,025	0,014	0,012	0,019	0,008	0,037	0,019	0,012	0,003	0,017	0,964
Ativ. Embed. Contínuos	0,017	0,008	0,009	0,010	0,042	0,021	0,005	0,012	0,035	0,006	0,014	0,008	0,008	0,979
Número Heads	0,010	0,003	0,003	0,029	0,006	0,005	0,002	0,010	0,018	0,026	0,006	0,001	0,013	0,989

Tabela 83 – Hiperparâmetros mais Relevantes para a Métrica no SAINT

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,710	0,823	0,305	0,159	0,392	0,055	0,222	0,224	-	0,048	0,082	0,046	0,057	0,260
Taxa Aprendizado	0,014	0,010	0,242	0,129	0,234	0,440	0,523	0,021	-	0,018	0,031	0,052	0,454	0,441
Dropout MLP	0,102	0,042	0,074	0,233	0,022	0,028	0,099	0,298	-	0,157	0,290	0,025	0,053	0,559
Decréscimo Taxa	0,022	0,013	0,028	0,051	0,091	0,019	0,017	0,065	-	0,019	0,075	0,679	0,043	0,653
Dropout Embed.	0,021	0,014	0,016	0,120	0,010	0,293	0,006	0,053	-	0,198	0	0,074	0,026	0,728
Dropout FF	0,047	0,009	0,022	0,067	0,052	0,022	0,012	0,062	-	0,242	0,071	0,051	0,051	0,787
Ativação Transformer	0,010	0,038	0,009	0,070	0,045	0,014	0,017	0,059	-	0,053	0,035	0,012	0,025	0,819
Dropout Atenção	0,008	0,023	0,031	0,022	0,018	0,008	0,014	0,054	-	0,068	0,098	0,008	0,016	0,850
Ativ. Embed. Contínuos	0,007	0,004	0,077	0,042	0,007	0,016	0,019	0,049	-	0,040	0,068	0,006	0,028	0,880
Ativação MLP	0,028	0,003	0,023	0,009	0,006	0,024	0,026	0,019	-	0,015	0,055	0,003	0,082	0,904
BatchNorm MLP	0,001	0,003	0,003	0,020	0,032	0,012	0,001	0,016	-	0,056	0,073	0,031	0,002	0,925
Número Heads	0,002	0,004	0,066	0,018	0,007	0,009	0,005	0,015	-	0,025	0,062	0,001	0,030	0,945

Tabela 84 – Hiperparâmetros mais Relevantes para o Tempo no SAINT

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Decréscimo Taxa	0,420	0,064	0,082	0,083	0,081	0,037	0,331	0,226	-	0,159	0,167	0,226	0,235	0,176
Dimensão Modelo	0,026	0,021	0,451	0,040	0,063	0,672	0,006	0,003	-	0,119	-	-	0,005	0,317
Batch	0,004	0,016	0,017	0,165	0,150	0,004	0,533	0,127	-	0,020	0,004	0,119	0,412	0,447
Número Blocos	0,069	0,033	0,068	0,038	0,310	0,122	0,006	0,226	-	0,116	0,200	0,082	0,020	0,555
Otimizador	0,056	0,427	0,019	0,122	0,014	0,012	0,012	0,043	-	0,034	0,026	0,112	0,062	0,633
Dropout Atenção	0,126	0,024	0,072	0,151	0,012	0,008	0,005	0,147	-	0,037	0,021	0,051	0,011	0,689
Dropout Embed.	0,177	0,029	0,041	0,046	0,020	0,042	0,008	0,014	-	0,125	0	0,046	0,042	0,742
Ativação Transformer	0,021	0,046	0,043	0,078	0,108	0,012	0,025	0,008	-	0,022	0,071	0,075	0,070	0,790
Dropout MLP	0,029	0,085	0,018	0,022	0,134	0,013	0,031	0,016	-	0,018	0,129	0,036	0,035	0,838
Dropout FF	0,010	0,022	0,034	0,068	0,025	0,009	0,008	0,108	-	0,088	0,024	0,058	0,061	0,880
Número Heads	0,011	0,009	0,012	0,037	0,002	0,005	0,015	0,009	-	0,010	0,278	0,006	0,007	0,914
Ativ. Embed. Contínuos	0,011	0,017	0,038	0,057	0,012	0,019	0,004	0,023	-	0,052	0,014	0,058	0,012	0,940
Ativação MLP	0,004	0,039	0,009	0,012	0,019	0,016	0,006	0,015	-	0,011	0,015	0,075	0,005	0,959

Tabela 85 – Hiperparâmetros mais Relevantes para a Métrica no FastFormer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Otimizador	0,591	0,657	0,449	0,437	0,703	0,022	0,314	0,431	0,156	0,052	0,075	0,413	0,201	0,346
Dropout MLP	0,047	0,033	0,043	0,067	0,040	0,064	0,222	0,054	0,131	0,051	0,402	0,158	0,041	0,451
Taxa Aprendizado	0,090	0,050	0,019	0,271	0,010	0,012	0,102	0,018	0,067	0,375	0,013	0,079	0,057	0,540
Decréscimo Taxa	0,039	0,073	0,135	0,064	0,016	0,039	0,095	0,082	0,023	0,055	0,090	0,060	0,021	0,601
Dropout FF	0,010	0,044	0,024	0,068	0,047	0,139	0,025	0,017	0,061	0,162	0,011	0,012	0,041	0,652
Ativ. Embed. Contínuos	0,013	0,018	0,054	0,010	0,017	0,032	0,022	0,044	0,115	0,046	0,129	0,033	0,077	0,699
Número Blocos	0,010	0,003	0,012	0,003	0,007	0,161	0,018	0,055	0,021	0,020	0,017	0,011	0,212	0,741
Ativação MLP	0,014	0,027	0,051	0,015	0,021	0,054	0,008	0,034	0,188	0,031	0,034	0,016	0,042	0,782
Dropout Atenção	0,085	0,017	0,033	0,014	0,052	0,076	0,055	0,011	0,032	0,037	0,035	0,031	0,042	0,822
Dropout Embed.	0,057	0,008	0,035	0,009	0,012	0,123	0,028	0,024	0,034	0,011	0	0,053	0,028	0,858
Ativação Transformer	0,017	0,016	0,013	0,004	0,009	0,053	0,038	0,043	0,059	0,050	0,037	0,023	0,056	0,890
Número Heads	0,002	0,009	0,014	0,009	0,005	0,023	0,015	0,029	0,053	0,018	0,029	0,010	0,102	0,914
BatchNorm	0,004	0,007	0,014	0,008	0,001	0,114	0,002	0,094	0,004	0,004	0,004	0,010	0,005	0,935
Última Camada Batch	0,003	0,011	0,063	0,007	0,016	0,024	0,011	0,017	0,003	0,012	0,013	0,020	0,009	0,951

Tabela 86 – Hiperparâmetros mais Relevantes para o Tempo no FastFormer

Hiperparâmetro	Bioconcentração	Concreto	Cardiologia	Abalone	Vinho	Banco	Feijão	Pulsar	Supercondutor	Múon	Diamante	SUS	Higgs	Média Acumulada
Decréscimo Taxa	0,375	0,461	0,337	0,391	0,241	0,614	0,272	0,452	0,397	0,602	0,682	0,163	0,146	0,395
Batch	0,012	0,014	0,022	0,162	0,257	0,009	0,394	0,026	0,019	0,004	0,014	0,183	0,534	0,522
Dropout MLP	0,034	0,114	0,110	0,063	0,069	0,023	0,047	0,009	0,058	0,088	0,032	0,110	0,068	0,585
Dimensão Modelo	0,033	0,011	0,134	0,004	0,020	0,181	0,128	0,013	0	0,076	-	-	0,004	0,646
Otimizador	0,032	0,035	0,031	0,018	0,026	0,015	0,008	0,097	0,203	0,029	0,092	0,041	0,027	0,696
Dropout Embed.	0,024	0,062	0,100	0,068	0,057	0,009	0,018	0,056	0,027	0,014	0	0,105	0,023	0,743
Dropout FF	0,050	0,062	0,037	0,114	0,104	0,028	0,007	0,022	0,037	0,052	0,015	0,040	0,041	0,790
Dropout Atenção	0,032	0,057	0,056	0,060	0,036	0,033	0,027	0,042	0,037	0,070	0,047	0,012	0,013	0,830
Ativação Transformer	0,071	0,029	0,018	0,016	0,011	0,037	0,029	0,023	0,036	0,011	0,025	0,031	0,033	0,858
Ativ. Embed. Contínuos	0,030	0,035	0,039	0,023	0,013	0,003	0,014	0,071	0,025	0,011	0,048	0,033	0,020	0,887
Taxa Aprendizado	0,044	0,056	0,011	0,005	0,009	0,014	0,004	0,007	0,018	0,005	0,006	0,115	0,038	0,912
Número Blocos	0,052	0,027	0,021	0,020	0,028	0,001	0,006	0,048	0,022	0,004	0,006	0,030	0,006	0,933
Ativação MLP	0,069	0,010	0,012	0,020	0,016	0,017	0,006	0,015	0,043	0,010	0,005	0,011	0,018	0,952
Número Heads	0,059	0,006	0,008	0,005	0,011	0,001	0,003	0,064	0,007	0,004	0,004	0,003	0,005	0,966

É possível notar nas tabelas 75, 77, 79, 81, 83 e 85 que para todas as redes neurais demonstradas a escolha do otimizador foi extremamente relevante para os resultados nos bancos de dados estudados, repetidamente superando a importância dada aos hiperparâmetros internos de cada modelo. De forma semelhante, nessas mesmas tabelas nota-se que a taxa de aprendizado

inicial e o decréscimo que ela sofre em cada época também repetidamente figuraram entre os hiperparâmetros mais relevantes para as mais diversas bases de dados.

Em linha com o concluído no artigo de Schmidt, Schneider e Hennig (2021), que demonstra uma pequena vantagem para os otimizadores baseados no Adam (KINGMA; BA, 2017) em diferentes *benchmarks*, foi percebido que o otimizador Adam, e sua variação AdamW (LOSHCHILOV; HUTTER, 2019), se mostraram como as melhores escolhas para a maior parte das redes neurais mais complexas.

Como demonstrado na tabela 87, das 77 combinações de hiperparâmetros escolhidas o Adam foi o melhor em 30 delas, o AdamW em 43, o SGD em 3, e o Adadelta (ZEILER, 2012) em somente uma.

Tabela 87 – Otimizador Escolhido na Melhor Combinação de Hiperparâmetros

Dataset	ResNet	TabNet	TabTransformer	FT-Transformer	SAINT	FastFormer
Bioconcentração	AdamW	Adam	AdamW	AdamW	AdamW	AdamW
Concreto	AdamW	AdamW	SGD	Adam	SGD	Adam
Cardiotocografia	AdamW	Adam	AdamW	Adam	Adam	AdamW
Abalone	AdamW	AdamW	AdamW	Adam	SGD	Adam
Vinho	AdamW	Adam	Adam	Adam	AdamW	Adam
Banco	AdamW	AdamW	Adam	AdamW	AdamW	Adam
Feijão	AdamW	AdamW	Adam	AdamW	Adam	Adam
Pulsar	AdamW	AdamW	Adam	Adam	AdamW	Adam
Supercondutor	Adam	AdamW	AdamW	Adam	-	Adam
Múon	AdamW	Adam	AdamW	Adam	Adam	AdamW
Diamante	AdamW	AdamW	AdamW	AdamW	Adadelta	Adam
SUS	AdamW	Adam	AdamW	AdamW	AdamW	AdamW
Higgs	AdamW	AdamW	AdamW	Adam	Adam	AdamW

7 DISCUSSÃO E TRABALHOS FUTUROS

A presença de múltiplas variáveis categóricas, ou tratadas como, aumentou de forma noticiável o tempo de processamentos das redes neurais mais complexas, podendo tornar esses métodos menos interessantes para lidar com bancos de dados com essa característica.

Os métodos de *Gradient Boosted Trees* mostraram, em todas as situações, ter um excelente custo-benefício, com performances consistentemente entre as melhores calculadas, atingidas em poucos segundos mesmo sem a utilização de uma GPU, com exceção do banco de dados da cardiocografia em que o uso do *Ordered Boosting* no CatBoost aumentou de forma considerável o seu tempo de processamento.

O SVM teve uma boa performance nos bancos de dados menores. Seu custo computacional aumenta de forma aproximadamente quadrática com o número de amostras (BOTTOU; LIN, 2006), o tornando um dos modelos mais demorados nos bancos de dados maiores, inclusive sendo necessário o uso somente de *kernels* lineares, que possibilitam implementações mais eficientes, nos bancos de dados dos atrasos no SUS e detecção do Bóson de Higgs.

MLP e ResNet estiveram repetidamente entre as melhores métricas das redes neurais, a simplicidade de suas arquiteturas não foi impedimento para os bons resultados.

Dentre as redes neurais baseadas em *Transformers*, o FT-Transformer demonstrou uma vantagem na sua performance em relação as demais.

Também foi possível notar em que na maior parte dos bancos de dados houve uma grande variação nas importâncias que modelos de categorias diferentes, como os *Gradient Boosted Trees* e redes neurais, assinalaram para as variáveis explicativas, mesmo quando o resultado das métricas não diferiu tanto.

7.1 TRABALHOS FUTUROS

Em trabalhos futuros seria interessante o uso de outros conjuntos de dados, com diferentes tamanhos amostrais e quantidade de variáveis explicativas, podendo dar uma comparação mais ampla para o desempenho dos algoritmos.

Nesse trabalho, para todas as redes neurais foi realizada a rotina de aprendizado em que a cada época a taxa de aprendizado é igual a da época anterior multiplicada por um valor entre 0 e 1, permitindo que os modelos possam atravessar grandes extensões da superfície de perda inicialmente, mas que consigam realizar um ajuste mais fino próximo do ponto de ótimo. Mas existem outras rotinas possíveis, como o *Cyclic Learning*, onde a taxa pode aumentar ou decrescer de acordo com a época, ou as baseadas em *plateaus*, onde a taxa só é reduzida quando a função de perda deixa de apresentar alguma melhoria significativa.

Há também várias outras técnicas propostas na literatura que poderiam ser aplicadas nas redes neurais, como o uso de *pretraining* e *data augmentation*.

Dada a alta importância atribuída aos otimizadores para os resultados das redes neurais, seria interessante que os vários outros propostos na literatura possam ser testados, avaliando

tanto as métricas obtidas quanto a capacidade de generalização.

Para o banco de dados de atrasos no SUS poderia ser incorporada uma análise espacial na predição, avaliando, por exemplo, a área atendida pelo posto e a densidade populacional, que pode levar pacientes que moram em distâncias maiores a serem dependentes de transporte público, podendo os tornar mais propensos ao atraso do que pacientes que moram em distâncias onde é viável a locomoção à pé até o local da consulta.

Para ele, sua análise também poderia ser beneficiada através do acompanhamento dos atrasos por um maior período de tempo, podendo inferir sobre efeitos temporais como a possível influência dos meses na variável resposta, o efeito de interação entre o dia e o mês, por exemplo com a proximidade da data da consulta a um feriado; além da incorporação de informações sobre se houve atrasos em consultas anteriores do mesmo paciente.

Outras variáveis interessantes seriam informações sobre se houve interrupção do transporte público no dia da consulta, além de questões climáticas da data como se está chovendo no dia o que pode, por exemplo, acarretar em enchentes na região.

Para o banco de dados dos supercondutores seria interessante a adição de outras informações dos experimentos, como a pressão que a temperatura crítica foi medida, e outras características do material, como sua estrutura cristalina.

Da mesma forma, para o banco de dados do concreto a análise seria beneficiada com mais informações das substâncias adicionadas, como através de uma análise físico-química da água e da granulometria e composição química da escória de alto-forno e das cinzas volantes.

Para o banco de dados dos feijões, novas variáveis que poderiam se provar importantes seriam fatores relacionados à coloração de cada grão.

Para o banco de dados dos vinhos há uma grande gama de possibilidades para estender a análise, avaliando informações relacionadas à turbidez, tempo de envelhecimento, presença de fenóis e flavonoides, maior detalhamento da concentração de outros ácidos orgânicos além do cítrico e tartárico, como o málico; o cultivar das uvas, se há a presença de mais de um tipo de uva em seu preparo. Além de dados da microrregião produtora, como sua distribuição geográfica, fenômenos climáticos que ocorreram durante a época de plantio, e seu *terroir*.

REFERÊNCIAS

- AGRAWAL, S. *Diamonds*. 2017. <<https://www.kaggle.com/shivam2503/diamonds>>. 43
- AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2019. 50
- AMARI, S. et al. Asymptotic statistical theory of overtraining and cross-validation. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 8, p. 985–996, 1995. 46
- ARIK, S. O.; PFISTER, T. *TabNet: Attentive Interpretable Tabular Learning*. 2020. 17, 18, 29, 30, 40
- ATLAS Collaboration. *Dataset from the ATLAS Higgs Boson Machine Learning Challenge 2014*. CERN Open Data Portal, 2014. Disponível em: <<http://opendata.cern.ch/record/328>>. 45
- BOTTOU, L.; LIN, C. jen. *Support Vector Machine Solvers*. 2006. 119
- CAMPOS, D. A. de et al. Sisporto 2.0: A program for automated analysis of cardiocograms. *The Journal of Maternal-Fetal Medicine*, Wiley, v. 9, n. 5, p. 311–318, 2000. Disponível em: <[https://doi.org/10.1002/1520-6661\(200009/10\)9:5<311::aid-mfm12>3.0.co;2-9](https://doi.org/10.1002/1520-6661(200009/10)9:5<311::aid-mfm12>3.0.co;2-9)>. 44
- CHEN, T.; GUESTRIN, C. Xgboost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Aug 2016. Disponível em: <<http://dx.doi.org/10.1145/2939672.2939785>>. 18
- CHOLLET, F. *Deep Learning with Python*. 2. ed. [S.l.]: Manning, 2021. ISBN 9781617296864. 29, 30, 31
- CHUI, M. et al. *Notes from the AI frontier: Applications and value of deep learning*. [S.l.], 2018. 30 p. 17
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. 18
- CORTEZ, P. et al. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, v. 47, n. 4, p. 547–553, 1998. 44
- DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, MIT Press - Journals, v. 10, n. 7, p. 1895–1923, out. 1998. Disponível em: <<https://doi.org/10.1162/089976698300017197>>. 42
- EDWARDS, A. L. Note on the “correction for continuity” in testing the significance of the difference between correlated proportions. *Psychometrika*, Springer Science and Business Media LLC, v. 13, n. 3, p. 185–187, set. 1948. Disponível em: <<https://doi.org/10.1007/bf02289261>>. 42

- FIX, E.; HODGES, J. L. *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*. Randolph Field, Texas: [s.n.], 1951. Disponível em: <<https://apps.dtic.mil/dtic/tr/fulltext/u2/a800276.pdf>>. 18, 21
- GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Mach. Learn.*, Kluwer Academic Publishers, USA, v. 63, n. 1, p. 3–42, apr 2006. ISSN 0885-6125. Disponível em: <<https://doi.org/10.1007/s10994-006-6226-1>>. 18, 23
- GORISHNIY, Y. et al. *Revisiting Deep Learning Models for Tabular Data*. 2021. 17, 18, 30, 35, 36
- GRISONI, F. et al. Investigating the mechanisms of bioconcentration through QSAR classification trees. *Environment International*, Elsevier BV, v. 88, p. 198–205, mar. 2016. Disponível em: <<https://doi.org/10.1016/j.envint.2015.12.024>>. 44
- GRISONI, F. et al. QSAR models for bioconcentration: Is the increase in the complexity justified by more accurate predictions? *Chemosphere*, Elsevier BV, v. 127, p. 171–179, maio 2015. Disponível em: <<https://doi.org/10.1016/j.chemosphere.2015.01.047>>. 44
- GÉRON, A. *Hands-on machine learning with scikit-learn and TensorFlow*. Sebastopol, CA: O'Reilly Media, 2017. 22, 23, 24, 25, 26, 27, 28
- HAMIDIEH, K. A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, Elsevier BV, v. 154, p. 346–354, nov. 2018. Disponível em: <<https://doi.org/10.1016/j.commatsci.2018.07.052>>. 43
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. 1. ed. New York, NY, USA: Springer New York Inc., 2001. (Springer Series in Statistics). 19, 21, 22
- HASTIE, T.; TIBSHIRANI, R.; TIBSHIRANI, R. J. Extended comparisons of best subset selection, forward stepwise selection, and the lasso. 2017. 20
- HE, K. et al. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770–778, 2016. 18, 29
- HO, T. K. Random decision forests. In: IEEE. *Proceedings of 3rd international conference on document analysis and recognition*. [S.l.], 1995. v. 1, p. 278–282. 18
- HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, v. 12, p. 55–67, 1970. 19
- HOPPEN, J. *Medical Appointment No Shows*. 2017. <<https://www.kaggle.com/joniarroba/noshowappointments>>. Accessed: 2020-11-05. 43
- HUANG, X. et al. *TabTransformer: Tabular Data Modeling Using Contextual Embeddings*. 2020. 17, 18, 30, 34, 35
- HUTTER, F.; HOOS, H.; LEYTON-BROWN, K. An efficient approach for assessing hyperparameter importance. In: XING, E. P.; JEBARA, T. (Ed.). *Proceedings of the 31st International Conference on Machine Learning*. Beijing, China: PMLR, 2014. (Proceedings of Machine Learning Research, 1), p. 754–762. Disponível em: <<https://proceedings.mlr.press/v32/hutter14.html>>. 114

- JAMES, G. et al. *An Introduction to Statistical Learning: with Applications in R*. 1. ed. [S.l.]: Springer, 2013. 19, 21, 24, 25, 26
- KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, v. 30, p. 3146–3154, 2017. 18
- KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2017. 118
- KOKLU, M.; OZKAN, I. A. Multiclass classification of dry beans using computer vision and machine learning techniques. *Comput. Electron. Agric.*, v. 174, p. 105507, 2020. 44, 77
- LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. In: *ICLR*. [S.l.: s.n.], 2019. 118
- LUNDBERG, S.; LEE, S.-I. *A Unified Approach to Interpreting Model Predictions*. 2017. 39, 40
- LYON, R. J. et al. Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Monthly Notices of the Royal Astronomical Society*, Oxford University Press (OUP), v. 459, n. 1, p. 1104–1123, abr. 2016. Disponível em: <<https://doi.org/10.1093/mnras/stw656>>. 43
- MCCAULEY, T. *Datasets derived from the Run2011A SingleElectron, SingleMu, DoubleElectron, and DoubleMu primary datasets*. CERN Open Data Portal, 2018. Disponível em: <<http://opendata.cern.ch/record/545>>. 43
- MCCULLOCH, W.; PITTS, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 127–147, 1943. 26
- MEINSHAUSEN, N. Relaxed lasso. *Computational Statistics and Data Analysis*, p. 374–393, 2007. 20
- MOLNAR, C. *Interpretable Machine Learning: A guide for making black box models explainable*. 2. ed. [s.n.], 2022. Disponível em: <christophm.github.io/interpretable-ml-book/>. 39
- MORO, S.; CORTEZ, P.; RITA, P. A data-driven approach to predict the success of bank telemarketing. *Decis. Support Syst.*, v. 62, p. 22–31, 2014. 44
- NASH, W. et al. The population biology of abalone (haliotis species) in tasmania. i. blacklip abalone (h. rubra) from the north coast and islands of bass strait. *Sea Fisheries Division, Technical Report No.*, v. 48, 01 1994. 44
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. 49
- PROKHORENKOVA, L. et al. *CatBoost: unbiased boosting with categorical features*. 2019. 18, 24
- RASCHKA, S. *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. 2020. 41, 42
- ROSENBLATT, F. *The perceptron - A perceiving and recognizing automaton*. Ithaca, New York, 1957. 27

- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, Springer Science and Business Media LLC, v. 323, n. 6088, p. 533–536, out. 1986. Disponível em: <<https://doi.org/10.1038/323533a0>>. 28
- SCHMIDT, R. M.; SCHNEIDER, F.; HENNIG, P. Descending through a crowded valley - benchmarking deep learning optimizers. In: *ICML*. [S.l.: s.n.], 2021. 118
- SOMEPELLI, G. et al. *SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training*. 2021. 17, 18, 30, 37, 38, 40
- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, v. 58, p. 267–288, 1994. 19
- VASWANI, A. et al. *Attention Is All You Need*. 2017. 30, 32, 33, 34, 37
- VIAL, G.; ESTERMANN, F. *Relaxed Lasso*. 2021. <<https://github.com/continental/RelaxedLasso>>. 49
- WU, C. et al. *Fastformer: Additive Attention Can Be All You Need*. 2021. 18, 30, 33, 36, 37
- YEH, I.-C. Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, Elsevier BV, v. 28, n. 12, p. 1797–1808, dez. 1998. Disponível em: <[https://doi.org/10.1016/s0008-8846\(98\)00165-3](https://doi.org/10.1016/s0008-8846(98)00165-3)>. 44
- YOU DEN, W. J. Index for rating diagnostic tests. *Cancer*, v. 3, n. 1, p. 32–35, 1950. 49
- ZAURIN, J. R.; MULINKA, P. *Pytorch-Widedeep*. 2021. <<https://github.com/jrzaurin/pytorch-widedeep>>. 49
- ZEILER, M. D. *ADADELTA: An Adaptive Learning Rate Method*. 2012. 118
- ZHANG, A. et al. *Dive into Deep Learning*. [S.l.: s.n.], 2021. <<https://d2l.ai>>. 29, 32
- ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 67, p. 301 – 320, 04 2005. 19, 20