



**UNIVERSIDADE ESTADUAL DE CAMPINAS**  
**Faculdade de Engenharia Elétrica e de Computação**

**CARLOS EDUARDO SOUZA MOREIRA**

**PREDIÇÃO DE FRETES AGRÍCOLAS BASEADO NA APLICAÇÃO DE  
TÉCNICAS DE APRENDIZADO DE MÁQUINA**

**CAMPINAS**

**2022**

CARLOS EDUARDO SOUZA MOREIRA

**PREDIÇÃO DE FRETES AGRÍCOLAS BASEADO NA APLICAÇÃO DE  
TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na área de Automação.

Supervisor/Orientador: Akebo Yamakami

Este trabalho corresponde à versão final da dissertação defendida pelo aluno Carlos Eduardo Souza Moreira, orientada pelo Prof. Dr. Akebo Yamakami

---

Assinatura do Orientador

CAMPINAS

2022

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

M813p Moreira, Carlos Eduardo Souza, 1979-  
Predição de fretes agrícolas baseado na aplicação de técnicas de aprendizado de máquina / Carlos Eduardo Souza Moreira. – Campinas, SP :[s.n.], 2022.

Orientador: Akebo Yamakami.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Logística. 2. Redes neurais (Computação)  
. 3. Aprendizado de máquina. 4. Máquina de vetores de suporte. 5. Transporte.

I. Yamakami, Akebo, 1947-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Predicting agricultural commodity transportation prices using machine learning algorithms

**Palavras-chave em inglês:**

Logistic  
Neural Network (Computing)  
Machine Learning  
Support vector machines  
Transportation

**Área de concentração:** Automação

**Titulação:** Mestre em Engenharia Elétrica

**Banca examinadora:**

Akebo Yamakami [Orientador]  
Priscila Cristina Berbert Rampazzo  
Romis Ribeiro de Faissol Attux

**Data de defesa:** 15-06-2022

**Programa de Pós-Graduação:** Engenharia Elétrica

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-8815-1289>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6197986895363789>

## COMISSÃO JULGADORA – DISSERTAÇÃO DE MESTRADO

Candidato: Carlos Eduardo Souza Moreira      RA: 099056

Data da defesa: 15 de junho de 2022

Título da Dissertação: “Predição de fretes agrícolas baseado na aplicação de técnicas de aprendizado de máquina”

Prof. Dr. Akebo Yamakami (Presidente)

Profa. Dra. Priscila Cristina Berbert Rampazzo

Prof. Dr. Romis Ribeiro de Faissol Attux

A Ata de Defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

## AGRADECIMENTOS

Esse mestrado foi um processo que teve início no segundo semestre do ano de 2009, momento em que fui aceito para cursar a disciplina de Linguagens Formais e Autômatos, ministrada pelo professor Arnaldo Vieira, no Instituto de Computação da Unicamp. Desde lá teve suas pausas, altos e baixos, tendo sido, em determinado momento, transferido para o Instituto de Matemática e Estatística da USP, e depois de volta para a Unicamp, dessa vez na FEEC. Dessa forma eu não tenho qualquer pretensão de tornar esse agradecimento abrangente o suficiente para tocar todos aqueles que contribuíram de alguma forma ao longo de todos esses anos, e por isso quero antecipadamente me desculpar caso o seu nome não esteja aqui.

Eu acho que o primeiro agradecimento vai para a pessoa que começou com isso tudo, Eduardo Fraga Moreira. Uma cabeça fantástica, enxadrista, fã de rock progressivo, técnico em contabilidade de formação, um homem que aprendeu a programar sozinho, a quem eu tive o privilégio de chamar de pai. Um homem sério e sensível, cujo o tempo endureceu, mas sem ter lhe retirado a ternura. Ele é o programador número zero dessa família, foi ele quem plantou uma semente binária na cabeça de um menino de 6 anos de idade. E foi com ele que eu aprendi que a lição mais preciosa que se possa passar adiante: que de todos os presentes que um pai pode dar a um filho, o mais precioso é o tempo.

Avançando na linha do tempo, tenho que agradecer a dois mestres que tive na graduação em Ciência da Computação, ainda no Rio Grande do Sul: a professora Ana Paula Lüdtke Ferreira, que de todos os professores que eu tive, foi a que mais disciplinas me ministrou, além de ter orientado meu trabalho de conclusão de curso. Foi ela quem em 1999 me pegou pela mão e, na condição de coordenadora do curso de Ciência da Computação, disse: “vamos fazer a sua troca de curso Cadu, você não é alguém de Sistemas de Informação, você é cientista da computação de coração”. O segundo trata-se do professor Ney Lemke, físico, criador de cachorros, e meu padrinho político. É de conhecimento de nós dois que o Ney me deu uma oportunidade que muda uma vida, ao me conceder uma bolsa de Iniciação Científica, e ter me doado um computador para que eu pudesse trabalhar em casa, sem nunca ter me cobrado nada. Mas o que ele talvez não saiba é que ele, sem perceber, me deu os fundamentos da consciência política que eu carrego até hoje.

Não posso deixar de mencionar o professor Akebo Yamakami, que concordou em ser orientador de alguém que ele nunca havia visto na vida, e quem vos escreve. Eu vejo isso como um ato de fé, tão raro nos dias de hoje, mas ao mesmo tempo tão essencial na nossa humanidade, presente apenas nos corações mais elevados.

Um agradecimento também ao professor Romis Attux, de quem, anos antes de ser seu aluno, eu havia ouvido falar nos quatro cantos da Universidade Estadual de Campinas, sempre de maneira elogiosa, respeitadora e ao mesmo tempo cientificamente excelente. O professor Romis entendeu algo que nem todos os mestres em sua posição compreendem: que adquirir conhecimento é essencial, mas que sem a humildade para passá-lo adiante, ele se torna tão valioso quanto um quilo de ouro em uma ilha deserta. Grande parte do conhecimento que eu tenho hoje na área de Aprendizado de Máquina, pilar desse trabalho, foi provocado por ele.

Vocês todos, de livre e espontânea vontade, ofereceram algo para que esse trabalho pudesse acontecer. Mas teve alguém que involuntariamente ofereceu algo diferente de conhecimento técnico, orientação cuidadosa, oportunidade e lições. Uma menina de 6 anos de idade (nessa data) que, sempre com um sorriso no rosto, abriu mão de manhãs, tardes, almoços, jantas, idas ao parque, passeios, e mesmo convívio com seu pai, para que ele pudesse terminar seu mestrado. E que mesmo sem entender o que ele estava fazendo na frente do computador durante tantas e longas horas, nunca deixou de apoiá-lo e oferecer a ele algo que era seu por direito. Filha, eu jamais vou poder devolver o período gasto que já passou, mas eu posso prometer que daqui para frente meu tempo será todo seu.

Por fim e (perdoem-me) mais importante, eu preciso agradecer à professora, pesquisadora, e minha esposa Andréa Leda Ramos de Oliveira. Vocês conhecem em maior ou menor grau a competência dela como profissional. Isso acontece porque a Andréa coloca amor em tudo o que ela faz. É por isso que ela é movida, e é através desse amor que ela encontra a disciplina e determinação através dos quais ela obteve todo o sucesso que conquistou até hoje. Essa mesma filosofia ela adotou no tratar desse trabalho, tendo se mantido como uma rocha, firme, forte e presente, mesmo nas piores tempestades. A Andréa acreditou nesse trabalho mesmo quando ninguém mais (nem eu) acreditava. E através do seu amor, alternando entre severidade e doçura, contribuiu de maneira fundamental para transformar uma ideia em uma realização, um insight num trabalho, um garoto num homem, e por isso eu tenho uma dívida impagável para com ela. Andréa, meu bem, quero

falar diretamente a você agora: você estava lá desde o começo, e estará no fim, muito tempo depois de tudo. Os especialistas dizem que o universo vai terminar em escuridão, e muitos tem medo disso. Mas eu não tenho, porque eu sei que você vai estar lá segurando a minha mão...como antes, como agora, como sempre.

Eu codifiquei, testei, organizei e escrevi esse trabalho. Mas isso não é tudo, e está muito longe de ser o suficiente para se chegar até aqui. Foram vocês, cada um de vocês, com suas contribuições que permitiram que tudo isso ocorresse. Vocês tocaram a minha vida e me trouxeram até esse momento. E agora eu estou aqui para dizer a vocês: muito obrigado.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e fez parte do projeto de pesquisa “Uma plataforma inteligente para previsão de preços de fretes agrícolas usando técnicas de mineração de dados” processo nº 2018/19571-1.

*“The Road goes ever on and on,  
Down from the door where it began.  
Now far ahead the Road has gone,  
And I must follow, if I can”*

(The Lord of the Rings, J. R. R. Tolkien)

## RESUMO

O custo de escoamento das safras agrícolas é um dos principais entraves à competitividade do Brasil no cenário internacional. O sistema rodoviário é o principal modal utilizado para o transporte de cargas agrícolas; e muitas vezes, a única opção para movimentação desse tipo de produto. Esta dependência pode levar a crises de desabastecimento em momentos de paralisações e greves. O impacto financeiro com o transporte sentido pelos agentes econômicos que atuam nas cadeias agrícolas poderia ser reduzido com a implementação de um sistema de gestão de fretes. Uma alternativa promissora para lidar com esse desafio é o uso de técnicas de mineração de dados, que são capazes de extrair padrões e tendências em grandes volumes de dados, razão pela qual têm sido cada vez mais utilizadas para dar suporte à decisão de gestores de diversas áreas, em detrimento à intuição, e aos métodos baseados em experiência. Isto posto, o objetivo deste trabalho é prever o preço de frete de commodities agrícolas através de modelos gerados a partir de Redes Neurais Multicamadas (MLP, Multi Layer Perceptron:) e Máquina de Vetores de Regressão (SVR, Support Vector Regression). A destacada participação do modal rodoviário ao longo do processo logístico de distribuição dos grãos faz com que a determinação dos custos de transporte relacionados a esta etapa seja uma importante ferramenta de gestão e tomada de decisão para os gestores do agronegócio. Os resultados indicaram que ambas as técnicas foram eficientes para estimar os preços fretes rodoviários de grãos agrícolas em termos de R-quadrado e raiz quadrada do erro quadrático médio (RMSE, Root Mean Squared Error). Na comparação das técnicas, a técnica SVR teve um desempenho superior, tendo alcançado uma R-quadrado de 0,8921 e RMSE de 8,0464 para o intervalo das rotas até 600km e 0,8924 e 21,1167 para as rotas mais longas que 600km, fazendo uso de modelos especializados em cada conjunto. Com o MLP as métricas foram inferiores, com R-quadrado 0,8785 e 8,5404 para rotas curtas e 0,8506 e 24,8818 para rotas longas também fazendo uso de modelos especializados. Em termos do uso do modelo geral ou com o particionamento do conjunto de dados em função da distância percorrida na rota de transporte, a utilização de dois modelos individuais especializados na partição do conjunto das rotas em duas faixas de até 600km ou superior, teve um melhor desempenho quando comparado a adoção de um modelo geral para todas as duas faixas de distâncias. Isso confirma a hipótese de que os preços dos fretes rodoviários apresentam comportamento distintos para diferentes faixas de distâncias.

Palavras-chave: Logística, Transporte, Máquina de Vetor de Suporte para Regressão, Rede Neural.

## ABSTRACT

The grain freight cost is one of the biggest obstacles for the Brazil's competitiveness in international scenario. The road freight still is majority adopted on agricultural products transporting, and not unusual, that is the only choice to carry them. The cost weight, felt by the agribusiness operators, which work with supply chains might be decreased after the adoption of a freight management system. A promising alternative to deal with this challenge is the use of data mining techniques, which are able to extract patterns and trends in large amounts of data, which is why they have been increasingly used to support management decision in different areas, in place of intuition and resolutions based on experience. Thus, the project major goal is to develop and implement an intelligent system for forecasting road freight prices for agricultural grains. Thus, the aim of this research is predicting the freight price of agricultural commodities using models generated from Multi Layer Perceptron (MLP) and Support Vector Regression (SVR). The massive use of freight road in the logistic and grain distribution, reinforces the importance of predicting freight prices, therefore a system that could provide that kind of information to agribusiness managers and decision makers, would be critical on their daily activities. The results indicated that both techniques were efficient in estimating agricultural grain road freight prices in terms of R-squared and root mean square error (RMSE). Comparing the techniques, the SVR technique outperformed achieving an R-squared of 0.8921 and RMSE of 8.0464 for the range of routes up to 600km and 0.8924 and 21.1167 for routes longer than 600km, adopting specialized models for each data partition. In the other hand MLP resulted an R-squared 0.8785 and 8.5404 RMSE for short routes and 0.8506 and 24.8818 for long routes also making use of specialized models. In terms of using the general model or with partitioning the dataset as a function of distance traveled on the transportation route, using two individual models specialized in partitioning the set of routes into two ranges of up to 600km or greater, performed better when compared to adopting a general model for all two ranges of distances. This confirms the hypothesis that road freight prices show different behavior for different distance ranges.

Key Words: Logistic, Transportation, Regression Support Vector Machines, Neural Network.

## LISTA DE ILUSTRAÇÕES

Figura 2.1 – Representação funcional de neurônio artificial.....	23
Figura 2.2 – Par de padrões linearmente separáveis.....	24
Figura 2.3 – Rede de <i>perceptrons</i> com uma camada intermediária .....	25
Figura 2.4 – Principais técnicas de otimização em Redes Neurais Artificiais .....	29
Figura 2.5 – Separação de duas classes. ....	31
Figura 2.6 – Duas classes de observação não separáveis. ....	34
Figura 2.7 – Interpretação geométrica da margem suave para SVM. ....	37
Figura 3.1 – Esquema do modelo de Predição de Fretes Rodoviários de Grãos.....	41
Figura 3.2 – Número de rotas por intervalo de preço de frete.....	43
Figura 3.3 – Diagrama UML de Atividade do processo de integração .....	45
Figura 3.4 – Diagrama UML de Atividade da etapa de aplicação da técnica .....	46
Figura 3.5 – Esquema de divisão do conjunto de dados em subconjuntos de treinamento, validação e teste.....	48
Figura 4.1 Box Plot do preço do frete da soja ao longo dos meses, 2012 - 2019.....	53
Figura 4.2 Box Plot do preço do frete da soja ao longo dos meses e período da safra, 2012 - 2019.....	53
Figura 4.3 – RMSE e $R^2$ para distância $\leq 600$ km.....	55
Figura 4.4 – RMSE e $R^2$ para distância $> 600$ km.....	56
Figura 4.5 – RMSE e $R^2$ para distância $\leq 600$ km com penalização L2. ....	58
Figura 4.6 – RMSE e $R^2$ para distância $> 600$ km com penalização L2. ....	59
Figura 4.7 – Matriz de correlação de correlação de Pearson.....	60
Figura 4.8 – Partial Dependence Plot para o Preço do Frete previsto pelo modelo. ....	61
Figura 4.9 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização <i>Ridge Regression</i> igual a 0,1.....	62
Figura 4.10 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização <i>Ridge Regression</i> igual a 100,0. ....	63
Figura 4.11 – RMSE, $R^2$ e número ideal de neurônios do conjunto distância $\leq 600$ km, com <i>Hold-out</i> e modelo dedicado.....	64
Figura 4.12 – RMSE, $R^2$ e número ideal de neurônios do conjunto distância $> 600$ km, com <i>Hold-out</i> e modelo dedicado.....	65
Figura 4.13 – RMSE e $R^2$ do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2. ....	67
Figura 4.14 – RMSE e $R^2$ do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2. ....	68
Figura 4.15 – Dispersão dos valores previstos pelos modelos individuais para rotas curtas e longas, e dos valores reais, com regularização <i>Ridge Regression</i> iguais a 0,1 e 100,0 respectivamente. ....	70
Figura 4.16 – RMSE e $R^2$ do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição. ....	71

Figura 4.17 – RMSE e $R^2$ do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição. ....	72
Figura 4.18 – RMSE e $R^2$ do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.....	74
Figura 4.19 – RMSE e $R^2$ do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.....	75
Figura 4.20 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização <i>Ridge Regression</i> igual a 120,0. ....	77
Figura 4.21 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização <i>Ridge Regression</i> igual a 10,0. ....	78
Figura 4.22 – RMSE, $R^2$ e número ideal de neurônios do conjunto das rotas menores ou iguais a 600 km, obtidos pelo <i>Hold-out</i> , através de um modelo dedicado. ....	79
Figura 4.23 – RMSE, $R^2$ e número ideal de neurônios do conjunto das rotas maiores que 600 km, obtidos pelo <i>Hold-out</i> , através de um modelo dedicado. ....	80
Figura 4.24 – RMSE e $R^2$ do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada aplicada à otimização métricas das rotas dessa partição na busca pelo valor da penalização. ....	82
Figura 4.25 – RMSE e $R^2$ do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização. ....	83
Figura 4.26 – Dispersão dos valores previstos pelos modelos individuais para rotas curtas e longas, e dos valores reais, com penalizações iguais a 560,0 e 10,0 respectivamente..	84

## LISTA DE TABELAS

Tabela 2.1 – Alguns dos hiperparâmetros utilizados na definição da rede neural. ....	26
Tabela 4.1 – Resultados do <i>Hold-out</i> para busca do número de neurônios ideal.....	55
Tabela 4.2 – Resultados do <i>Hold-out</i> para busca da variância ideal .....	57
Tabela 4.3 – Resultados do <i>Hold-out</i> para busca do valor da penalização (L2) ideal.....	59
Tabela 4.4 – Resultados dos dois modelos de MLP única .....	61
Tabela 4.5 – Resultados do <i>Hold-out</i> para busca pela variância ideal .....	66
Tabela 4.6 – Resultados dos dois modelos de MLP independentes, aplicados sobre os conjuntos de treinamento e teste.....	69
Tabela 4.7 – Resultados do <i>Hold-out</i> aplicado na busca pelo valor de <i>Gamma</i> ideal.....	72
Tabela 4.8 – Resultados do <i>Hold-out</i> aplicado na busca pela variância ideal, para uso em PCA .....	73
Tabela 4.9 – Resultados do <i>Hold-out</i> aplicado na busca pelo valor da penalização (L2) ideal. ....	76
Tabela 4.10 – Resultados dos dois modelos de SVR única, aplicados sobre os conjuntos de treinamento e teste. ....	76
Tabela 4.11 – Resultados do <i>Hold-out</i> aplicado na busca pela variância ideal, para uso em PCA .....	81
Tabela 4.12 – Resultados dos dois modelos de SVR independentes, aplicados sobre os conjuntos de treinamento e teste.....	84
Tabela 4.13 – Resultados obtidos de SVR e MLP a partir do conjunto de teste.....	86

## SUMÁRIO

1	INTRODUÇÃO.....	16
2	FUNDAMENTAÇÃO TEÓRICA .....	19
2.1	Dinâmica do mercado de fretes agrícolas .....	19
2.2	<i>Multi Layer Perceptron</i> (MLP): alguns apontamentos.....	23
2.2.1	Algoritmo de Retropropagação do Erro .....	28
2.3	Support Vector Machine (SVM) e Support Vector Regression (SVR): alguns apontamentos.....	31
2.3.1	Margem Suave.....	37
2.3.2	SVR – A Versão para Regressão do SVM .....	39
3	METODOLOGIA.....	41
3.1	Descrição dos dados.....	41
3.2	Tratamento dos dados .....	42
3.3	Aplicação da Técnica e Refinamento .....	46
3.3.1	Escolha das Métricas .....	47
3.3.2	Validação Cruzada.....	48
3.3.3	Análise de Componentes Principais .....	49
3.3.4	Remoção de Fertilizantes.....	49
3.3.5	Particionamento do Conjunto de Dados .....	50
3.3.6	Regularização .....	50
4	RESULTADOS E DISCUSSÕES.....	52
4.1	Resultados da Aplicação da <i>MLP</i> .....	53
4.1.1	MLP Única – Um Modelo para Duas Partições .....	54
4.1.2	MLPs Individuais – Dois Modelos para Duas Partições .....	63
4.2	Resultados da Aplicação da <i>SVR</i> .....	70
4.2.1	SVR Única – Um Modelo para Duas Partições.....	70
4.2.2	SVRs Individuais – Dois Modelos para Duas Partições .....	78
4.3	MLP vs SVR.....	85
5	CONCLUSÕES .....	89
6	REFERÊNCIAS .....	91
	APÊNDICE .....	96

## 1 INTRODUÇÃO

Com dimensões continentais e grandes extensões de terras férteis, o Brasil desempenha um papel importante no mercado agrícola internacional. Principalmente sob os avanços tecnológicos, que viabilizaram progressos significativos na produção das commodities. Porém, para que estas melhorias sejam de fato eficazes, a superação dos entraves relacionados à logística de distribuição devem superados (OLIVEIRA et al., 2022).

O Brasil ocupa posição de destaque no cenário internacional de commodities agrícolas, com destaque para soja e milho. O Brasil é o maior produtor mundial de soja e quarto em milho, responsável por 38% da produção mundial de soja com uma produção estimada em 144 milhões de toneladas na safra 2021/22 (USDA, 2022). Quanto às exportações, o Brasil é o maior exportador de soja, 94 milhões de toneladas na safra 2021/22, e o terceiro maior exportador de milho, 30 milhões de toneladas na safra 2020/21 (USDA, 2022).

Apesar da importância para economia nacional, estas *commodities* perdem competitividade frente a outros países produtores por conta das deficiências relacionadas à logística de escoamento da produção (BARTHOLOMEU et al., 2019). O custo de escoamento das safras agrícolas é um dos principais entraves à competitividade do Brasil, que não consegue transformar as vantagens dos ganhos de produtividade do campo em competitividade na comercialização, principalmente em função dos altos custos de transporte (ROULET; CAIXETA FILHO, 2016).

As estratégias dos setores público e privado têm permeado a agenda do agronegócio para a superação de entraves relacionados ao custo Brasil (OLIVEIRA, 2014). Em especial para o agronegócio, as várias proposições sempre indicam a necessidade de reduzir os custos de transportes para aumentar a eficiência das cadeias agroindustriais.

Uma parcela significativa dos ganhos da produção agrícola brasileira é perdida com os altos custos de escoamento (EMBRAPA, 2016). O setor agrícola não consegue transformar as vantagens dos ganhos de produtividade do campo em competitividade na comercialização, principalmente em função dos altos custos de transporte (ROULET; CAIXETA FILHO, 2016). Segundo Correa e Ramos (2010), cerca de 25% da receita de vendas da produção de soja estão comprometidas com os custos de transporte.

A participação do transporte intermodal na movimentação de cargas não está alinhada com a grande extensão territorial brasileira. No Brasil, a matriz de transporte de cargas é altamente concentrada no modal rodoviário responsável por 64,9% da movimentação (CNT, 2022). Mesmo possuindo a maior extensão rodoviária da América Latina, 1,72 milhões de quilômetros de rodovias, apenas 12,5% são pavimentadas (CNT, 2022).

O sistema rodoviário é o principal modal utilizado para o transporte de cargas agrícolas; e muitas vezes, a única opção para movimentação desse tipo de produto, devido à pouca disponibilidade de ferrovias e à baixa navegabilidade comercial das hidrovias (OLIVEIRA et al., 2021).

Importante destacar que o frete rodoviário acaba sendo o balizador dos fretes dos outros modais; ou seja, é um indicador do mercado de fretes, tanto para o ferroviário quanto para o hidroviário. Os operadores dos demais modais atrelam os seus preços ao frete praticado pelo segmento rodoviário (OLIVEIRA et al., 2013).

Um desafio que se coloca para gestores da cadeia da soja, além de elaborar estratégias comerciais, é o de administrar a sazonalidade da produção dos grãos e a influência sobre os fretes (KAVUSSANOS; ALIZADEH-M, 2001). No período de colheita, o aumento da demanda por transporte até os portos leva à uma elevação dos fretes rodoviários, estimulando que as empresas ferroviárias e hidroviárias também aumentem seus preços (OLIVEIRA, 2006). Como os fretes ferroviários e hidroviários estão indexados ao comportamento do frete rodoviário, no período de colheita da safra os custos de transporte aumentam de forma geral.

Diante destas flutuações sazonais de preço os agentes do setor podem planejar o transporte e buscar a antecipação da contratação dos operadores logísticos (OLIVEIRA, 2006) e/ou firmar contratos de longo prazo com os fretadores (KAVUSSANOS; ALIZADEH-M, 2001). Independente da estratégia adota é preciso ampliar a previsão das despesas com transporte, sobretudo o rodoviário.

O uso de modelos preditivos já foi aplicado para o cálculo de custos totais de uma rede de transporte Janic (2007). Para previsão do frete rodoviário, trabalhos anteriores fazem-no mais comumente através de modelos de regressão (FERREIRA, 2016; COPPE et al., 2020), simulação de cenários (BLOIS; MARTINS, 2017; de BOK, et al., 2018) e modelo gravitacional (GÜLER, 2014).

Isto posto, o objetivo da pesquisa é prever o preço de frete de commodities agrícolas através de modelos gerados a partir de Redes Neurais Multicamadas (MLP, *Multi Layer Perceptron*) e Máquina de

Vetores de Regressão (SVR, *Support Vector Regression*). A hipótese que queremos verificar é se as técnicas de predição, quando submetidas à um conjunto de dados suficientemente representativos, são capazes de gerar modelos robustos para prever o comportamento do preço de frete de grãos agrícolas sob uma margem de erro aceitável.

O capítulo 2 irá apresentar um panorama geral da dinâmica do mercado de fretes e os principais fundamentos de Multi Layer Perceptron (MLP) e Support Vector Regression (SVR). O capítulo 3 irá apresentar a metodologia utilizada, a descrição dos dados e dos parâmetros adotados. O capítulo 4 apresentará os resultados obtidos e a comparação dos modelos preditivos a partir das duas técnicas propostas. E por fim, as conclusões deste trabalho serão descritas no capítulo 5.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os referenciais teóricos quanto ao estado da arte da formação do mercado de fretes agrícolas, *Multi Layer Perceptron* (MLP) e *Support Vector Regression* (SVR), direcionados às aplicações no que tange aos modelos preditivos.

### 2.1 Dinâmica do mercado de fretes agrícolas

O transporte rodoviário possui algumas vantagens em relação aos outros modais de transporte em função da sua maior disponibilidade, facilidade de contratação, capilaridade da malha e eficiência no transporte porta-a-porta (OLIVEIRA, 2014). No transporte com caminhões, as operações de carregamento são mais simples, as cargas podem ser combinadas se for necessário e entregue com agilidade, dando aos transportadores uma vantagem sobre outros modais de transporte ao atuarem no mercado de frete spot, sem a necessidade de contratos pré-estabelecidos (LIMA et al., 2016)

O mercado de frete rodoviário no Brasil, especialmente o de cargas agrícolas, não sofre nenhum tipo de mediação pelo governo quanto a política de preços, o que implica que os preços são formados a partir da dinâmica de oferta e demanda pelo serviço de transporte (ARAUJO et al., 2014; PÉRA et al., 2018;).

A formação do preço do transporte rodoviário, além dos custos fixos e variáveis da atividade, incorpora também fatores locais e conjunturais. Ortúzar e Willunsen (1995), Côrrea Jr. e Caixeta-Filho (2003), Lima et al., (2016), Cangussu et al. (2013), Oliveira et al. (2013) e Perú et al. (2018), analisaram as principais variáveis que têm influenciado o custo de transporte rodoviários de cargas agrícolas, com destaque para:

- i. distância percorrida e diferentes faixas de quilometragem;
- ii. especificidade da carga transportada;
- iii. volumes transportados;
- iv. sazonalidade da produção (períodos de safra e entressafra);
- v. peculiaridades regionais (rotas de transporte tradicionais);
- vi. possibilidade de carga de retorno (evitar o retorno vazio);
- vii. concorrência e/ou complementaridade com outras modalidades de transporte;
- viii. estado de conservação das vias;

- ix. pedágios e balanças ao longo das vias;
- x. risco da carga, cargas perigosas/produtos inflamáveis ou visados para roubo;
- xi. manuseio do produto (facilidade de se carregar e descarregar o veículo);
- xii. prazo de entrega.

Uma categorização das variáveis que interferem a formação do preço do frete de produtos agrícolas foi feita por Lima et al., (2016). Além da divisão de custos fixos (despesa com pessoal e com o veículo) e custos variáveis (combustível, lubrificação e manutenção), destacam aspectos como facilidade de manuseio e particularidades, condições de mercado, infraestrutura regional, competição e/ou sinergia com outros modais de transporte.

Em Perú et al. (2018) é possível verificar como o preço do frete rodoviário, de uma série de produtos agrícolas, para uma mesma faixa de distância apresenta diferenças entre os tipos de produtos, em função das suas características de mercado.

Os autores analisaram também a sazonalidade dos preços de fretes, que é influenciada por uma série de fatores, mas destacam a demanda de serviços de transporte. Isto porque, em épocas de colheita, particularmente em regiões de baixa capacidade de armazenagem, existe uma alta demanda por serviço de transporte para movimentar os produtos, implicando aumento do preço do frete.

O estudo de Oliveira (2006) já apontava interferência da armazenagem sobre o preço do frete rodoviário. O autor mostra que a concentração nos períodos de safra também exige maior capacidade dos portos de exportação de produtos agrícolas. Esse cenário, somado a uma relativa escassez de armazenagem, faz com que os operadores do mercado de *commodities* agrícolas tenham de planejar sua logística com cuidado, de maneira a reduzir os custos, aumentando assim sua competitividade.

Uma alternativa para o setor, na tentativa de equacionar a sazonalidade do mercado, é o planejamento e a contratação antecipada dos operadores logísticos. Embora a distância seja uma das variáveis que mais interfere na formação do preço do frete rodoviário dos produtos agrícolas, é importante destacar papel fundamental do mecanismo de negociação entre os agentes que atuam no mercado. A dinâmica competitiva deste mercado se dá através de uma grande disputa de forças. A demanda por transporte, na maioria dos casos na figura das tradings agrícolas, detém o poder de negociação frente à oferta de serviço de transporte. Isso porque as tradings agrícolas, em sua maioria empresas transnacionais, possuem grande volume de carga agrícola a ser escoada por um grande número

de pequenas empresas de transporte e caminhoneiros autônomos. Esta pulverização dos ofertantes de transporte traz vantagens às tradings agrícolas no momento da negociação dos fretes rodoviários (OLIVEIRA et al., 2013).

Para Araújo et al. (2014), os baixos fretes rodoviários praticados no mercado brasileiro acontecem pelo excesso de serviços de transportadores autônomos, em geral, contratados por grandes transportadoras. Por sua vez, as empresas transportadoras são contratadas por fabricantes de produtos para transportar matéria prima, ou no caso agroindustrial pelas tradings agrícolas, até a indústria ou mercados consumidores finais.

A distorção no mercado de frete rodoviário também se dá por conta do chamado círculo vicioso do transporte rodoviário de carga, com pequenas barreiras de entrada e elevadas barreiras de saída (WANKE, 2012). Isso porque existe uma facilidade para ingressar neste mercado (comprar ou arrendar um caminhão, inclusive com crédito subsidiado) e essa facilidade cria um aumento da oferta e, conseqüente, redução do preço do frete. Esse rebaixamento do nível de preço faz com que o pequeno empresário/autônomo tenha diversas dificuldades, dentre elas: falhas na manutenção dos caminhões, jornadas de trabalho excessivas, sobrecarga nos veículos, incapacidade de renovação de frota, aumento da idade média da frota, aumento do índice de acidentes e inadimplência fiscal. Assim, existe uma grande barreira para saída deste mercado e venda dos ativos/frota, retroalimentando o círculo vicioso.

Somam-se a esses fatores de formação de preço de frete rodoviário os movimentos de paralisação de caminhoneiros, o risco de desabastecimento e as discussões quanto ao estabelecimento de uma Tabela de Preços Mínimos, elementos que intensificam a discussão do modelo rodoviário vigente e a proposição da presente proposta de pesquisa de um modelo de previsão de preço de frete rodoviário como forma de ampliar as ferramentas de apoio para o setor agroindustrial.

Com já observado por Perú et al. (2018), o tabelamento mínimo de fretes causa impactos negativos à estrutura de mercado, aumento o custo logístico para o agronegócio brasileiro. Além disso, tal tabela pode desestimular a diferenciação do transporte em termos de melhoria da qualidade do serviço e, ainda levar a verticalização das operações de transporte pelas tradings agrícolas e empresas da agroindústria.

Quanto aos modelos preditivos para o frete rodoviário, trabalhos anteriores fazem-no mais comumente através do uso de técnicas estatísticas como regressão e séries temporais (OLIVEIRA;

YAMAKAMI, 2006, CANGUSSU et al., 2013, COSTA, 2014; FERREIRA, 2016; BLOIS; MARTINS, 2017; COPPE et al., 2020).

Acerca do uso de técnicas de aprendizado de máquina, esta é usual na previsão do preço de uma variedade de produtos e ativos financeiros, passando por imóveis, *commodities* diversas, e títulos de toda espécie. Abidoye et al. (2019) fez uso de Redes Neurais Artificiais, Máquinas de Vetor Suporte, e Modelo Auto-Regressivo de Médias Móveis, estabelecendo um comparativo entre elas, usando como estudo de caso a previsão do índice de preço de imóveis de Hong Kong. A motivação é a prevenção à perda de lucros e falência devido às inevitáveis altas e bolhas inerentes ao mercado. Como resultado os autores identificaram o melhor desempenho da técnica de Redes Neurais Artificiais sobre as demais.

No campo das *commodities*, Chiroma, et al. (2015) propuseram um modelo evolucionário que faz uso de Rede Neural Artificial e Algoritmo Genético (AG), para prever o preço do petróleo sob a cotação do índice *West Texas Intermediate* (WTI), de forma a oferecer suporte às decisões de gestores públicos e privados, no desenvolvimento de estratégias de curto e longo prazo. O modelo em si apresenta uma característica dinâmica devido ao uso do AG no ajuste simultâneo dos pesos, *bias*, e topologia da rede neural.

## 2.2 Multi Layer Perceptron (MLP): alguns apontamentos

Uma Rede Neural Artificial (RNA) é uma técnica de aprendizado de máquina baseada nas estruturas cerebrais. McCulloch e Pitts (1943), propuseram o primeiro modelo computacional do neurônio como uma unidade de processamento binária que apresentava elementos tais como: o limiar do disparo presente em cada neurônio, a estrutura estática da rede em função do tempo e o impedimento do disparo do neurônio através da ativação da entrada inibitória. Avançando até 1958, Frank Rosenblatt introduz o modelo de neurônio chamado *Perceptron* (Figura 2.1) (HAYKIN, 2008).

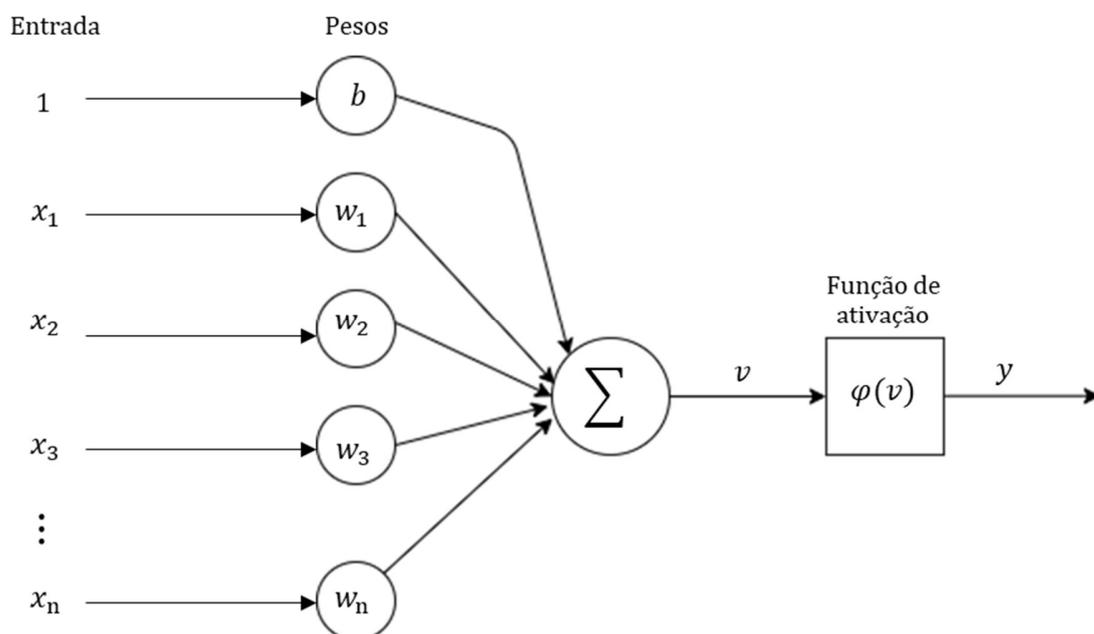


Figura 2.1 – Representação funcional de neurônio artificial

Fonte: Adaptado de Haykin (2008)

Neste tipo de rede neural, cada neurônio possui um conjunto de entradas  $x_1, \dots, x_n$  e cada entrada  $x_i$  passa a receber um peso ajustável  $w_i$ , pelo qual são multiplicadas e, então, mais uma entrada é adicionada, essa agora sem peso associado e denominada *bias* ( $b$ ), definido por  $w_0$ , com  $x_0 = 1$  (Figura 2.1), expressado por:

$$v = \sum_{i=0}^n w_i x_i = \mathbf{w}^T \mathbf{x} \quad (2.1)$$

O *Perceptron* de Rosenblatt é, essencialmente, uma rede com uma única camada. Em função desta característica, e para que possa operar de maneira eficaz, o conjunto de dados submetido a esta rede deve ser separável através de um hiperplano (Figura 2.2), ou uma reta, caso o conjunto de dados pertença a  $R^2$ . De fato, a separação linear é garantida pelo Teorema da Convergência do Perceptron (HAYKIN, 2008), que é a prova do algoritmo de mesmo nome.

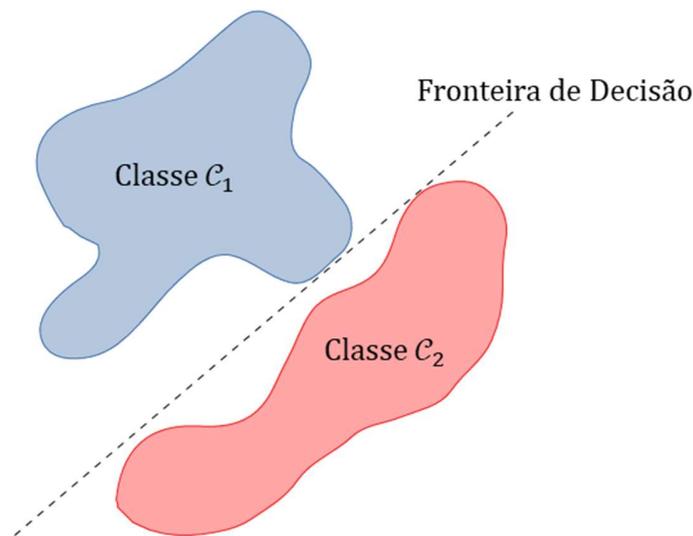


Figura 2.2 – Par de padrões linearmente separáveis

Fonte: Adaptado de Haykin (2008)

Todas as entradas são combinadas de maneira linear e, por fim, submetidas a uma função do tipo degrau, conhecida por *função de ativação*  $\varphi(v)$  que resulta na saída  $y$ , que executa um mapeamento para o domínio binário, permitindo dessa forma uma classificação em duas categorias.

$$y = \begin{cases} 1, & \text{se } v > 0 \text{ para toda entra } \mathbf{x} \text{ pertencente a classe } \mathcal{C}_1 \\ 0, & \text{se } v \leq 0 \text{ para toda entra } \mathbf{x} \text{ pertencente a classe } \mathcal{C}_2 \end{cases} \quad (2.2)$$

O próximo passo é o cálculo do erro  $e$ , que deve representar a diferença entre o resultado esperado e o resultado obtido, respectivamente  $d$  e  $y$ .

$$e = d - y \quad (2.3)$$

O erro juntamente com o hiperparâmetro da rede neural  $\eta$ , conhecido por taxa de aprendizado, é utilizado para recálculo dos pesos  $W$  associados às entradas.

$$w' = w + \eta e x \quad (2.4)$$

Ao final do processo de ajuste de todos pesos associados às entradas dos neurônios, temos uma etapa finalizada, denominada *época*.

Por outro lado, para que uma rede neural composta por neurônios do tipo *Perceptron* seja capaz de classificar padrões não linearmente separáveis, basta que pelo menos uma camada intermediária (ou oculta) seja adicionada à rede (Figura 2.3).

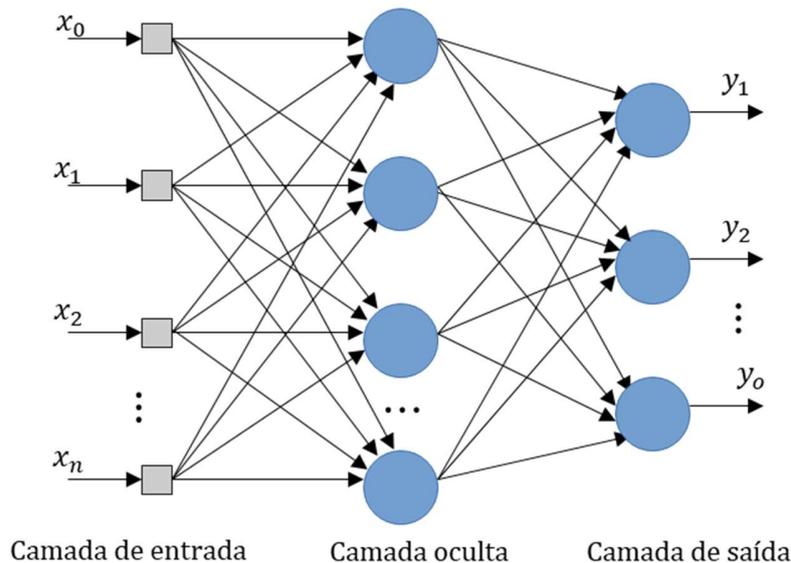


Figura 2.3 – Rede de *perceptrons* com uma camada intermediária

Fonte: Adaptado de Haykin (2008)

Os neurônios ocultos atuam como identificadores de padrões nos dados (ou simplesmente, *feature detectors*). Desta forma, eles desempenham um papel crucial na operação nessa rede de *perceptrons* multicamada (MLP). À medida que se dá o processo de aprendizado ao longo da MLP, os neurônios ocultos começam a gradualmente descobrir os padrões de dados que se sobressaem e, desta forma, caracterizam o conjunto de treinamento. Isso ocorre através de uma transformação não linear executada sobre os dados de entrada e que resulta em um novo espaço, denominado espaço de variáveis (*feature*

*space*). Nesse novo espaço, a tarefa de separação dos padrões em classes pode ser mais facilmente executada quando comparado ao espaço de dados de entrada original. É justamente a formação desse espaço de variáveis, através de um aprendizado supervisionado, que distingue o MLP do *Perceptron* de Rosenblatt, com uma única camada. Ainda, é importante ressaltar que a adição de uma única camada intermediária à rede já confere a capacidade de aproximação universal a ela, conforme mostrado por Cybenko (1989).

A MLP descrita na Figura 2.3 possui uma única camada escondida com  $n$  neurônios e, adicionalmente, com as características do comportamento da rede neural (Tabela 2.1), são algumas das possíveis maneiras de se alterar o comportamento da rede neural e que são denominadas hiperparâmetros.

Tabela 2.1 – Alguns dos hiperparâmetros utilizados na definição da rede neural.

Nome do Hiperparâmetro	Característica
Número de camadas escondidas	Diretamente proporcional à capacidade de representação do modelo.
Neurônios por camada	Da mesma forma que o número de camadas ocultas, esse hiperparâmetro está ligado àquilo que o modelo consegue representar.
Taxa de aprendizado	Uma taxa de aprendizado mal ajustada resulta em um modelo com baixa capacidade efetiva, devido às falhas de otimização.
Função de ativação	Função utilizada para mapear a composição das entradas. Algumas possibilidades são: função logística, tangente hiperbólica, e função retificadora.
Parada antecipada	Sinaliza a parada no processo de treinamento da rede quando a métrica selecionada não apresenta mais melhora significativa.

O ajuste manual dos valores para os hiperparâmetros demanda conhecimentos sobre o relacionamento entre eles, o erro de treinamento, o erro de generalização e os recursos computacionais demandados pelo algoritmo. O que significa estabelecer uma fundamentação sólida a respeito dos mecanismos internos da rede neural. Uma estratégia que demanda menos conhecimento, e ainda assim tem dado resultado para solução de problemas do mundo real (GOODFELLOW et al., 2016), é a adoção de técnicas de Validação Cruzada aplicadas a um conjunto de validação extraído a partir do conjunto original dos dados. Em uma divisão de, por exemplo, 80% (treinamento) e 20% (validação), o algoritmo selecionado para Validação Cruzada é exclusivamente aplicado sobre o segundo conjunto, e faz uso de

uma técnica de aprendizado de máquina, até que seja encontrado o hiperparâmetro que resulte no melhor valor de métrica dentre todas as iterações.

Um dos hiperparâmetros (Tabela 2.1) é a função de ativação. Já no modelo de McCulloch e Pitts (1943) a ideia dela estava presente na forma de um degrau, que transformava a composição das entradas do neurônio em valores booleanos. Mas foi no Perceptron que ela efetivamente ganhou sua forma atual de função.

O único requisito que a função deve atender para que possa ser utilizada como função de ativação pelo neurônio é a possibilidade de derivação (HAYKIN, 2008). Dois exemplos de funções de ativação não lineares diferenciáveis continuamente, comumente utilizadas em MLPs, são formas da sigmoide não linear: a função logística (equação 2.5), e a função tangente hiperbólica (equação 2.6)

$$\varphi(v) = \frac{1}{1 + e^{(-av)}} \quad a > 0 \quad (2.5)$$

e

$$\varphi(v) = a \tanh(b v) \quad (2.6)$$

O critério de parada antecipada (Tabela 2.1) se apresenta como uma tentativa de evitar longos, e potencialmente, infrutíferos períodos de treinamentos da rede.

Comumente uma MLP treinada através de um algoritmo de retropropagação de erro (HAYKIN, 2008) aprende em estágios, partindo de funções de mapeamentos simples, e progredindo na direção de funções mais complexas à medida que o treinamento avança. Esse processo é exemplificado pelo fato de que, em um cenário típico, o erro médio quadrático decresce com o passar das épocas do algoritmo, da seguinte forma: no início o erro é alto, sofrendo um caimento rápido, mas a seguir entra em uma fase em que o erro passa a diminuir mais lentamente, tão logo a rede trilha seu caminho em direção a um mínimo local na superfície de erro. Tendo por objetivo uma alta generalização, torna-se muito difícil determinar o melhor momento de interromper o treinamento da rede, usando como guia tão somente a curva de aprendizado. Por outro lado, a postergação da parada poderia resultar no sobre-ajuste do modelo (HAYKIN, 2008).

Uma estratégia para se evitar o sobre-ajuste é a utilização de uma técnica de Validação Cruzada a cada quantidade de épocas, com intuito de medir o erro. Para isso, faz-se necessária a divisão do conjunto de dados entre treinamento e validação. O treinamento ocorre naturalmente sobre o primeiro

conjunto, quando o número de épocas é atingido, o conjunto de validação deve ser avaliado e o erro resultante analisado. Por fim é dada continuidade ao treinamento novamente, até que se atinja o próximo limite de épocas. Esse processo é conhecido por *parada antecipada para método de treinamento* (HAYKIN, 2008).

Ainda assim o método descrito não garante que será atingido a generalização ótima, uma vez que a curva do erro de generalização pode apresentar uma série de mínimos locais, inviabilizando a exata determinação do momento ideal de parada (PRECHELT, 1998a). Uma investigação empírica em MLPs executada por Prechelt (1998b) demonstrou experimentalmente que de fato existe uma relação entre tempo de treinamento e performance da generalização. Baseado nos resultados obtidos a partir de 1.296 sessões de treinamento, 12 diferentes problemas, e 24 arquiteturas de redes distintas, conclui-se que, na presença de 2 ou mais mínimos locais, a opção por um critério de parada mais lento (ou seja, um critério que interrompeu o treinamento mais tardiamente quando em comparação com outros) resultou em um ganho de generalização de apenas 4% em média, ao custo de tempo de treinamento quatro vezes maior, em média.

### 2.2.1 Algoritmo de Retropropagação do Erro

O processo de aprendizado em uma rede neural é conduzido por um algoritmo de treinamento. O papel desse algoritmo é minimizar o índice de perda, que é dado por uma função que mede a performance da rede para o conjunto de dados submetido a ela. Dentre as diversas técnicas desenvolvidas para esse fim podemos mencionar as mais utilizadas: i. Gradiente descendente, ii. Gradiente conjugado, iii. Método de Newton, iv. Quasi Newton, v. Levenberg Marquardt (HAYKIN, 2008). Cada uma dessas técnicas apresenta requisitos de memória e velocidade, sendo a Busca no Gradiente Descendente a de menor consumo de memória, e menor velocidade, enquanto que a técnica Levenberg Marquardt detém o maior gasto de memória e maior velocidade. O comparativo de performance das cinco principais técnicas utilizadas em Redes Neurais Artificiais é apresentado na Figura 2.4.

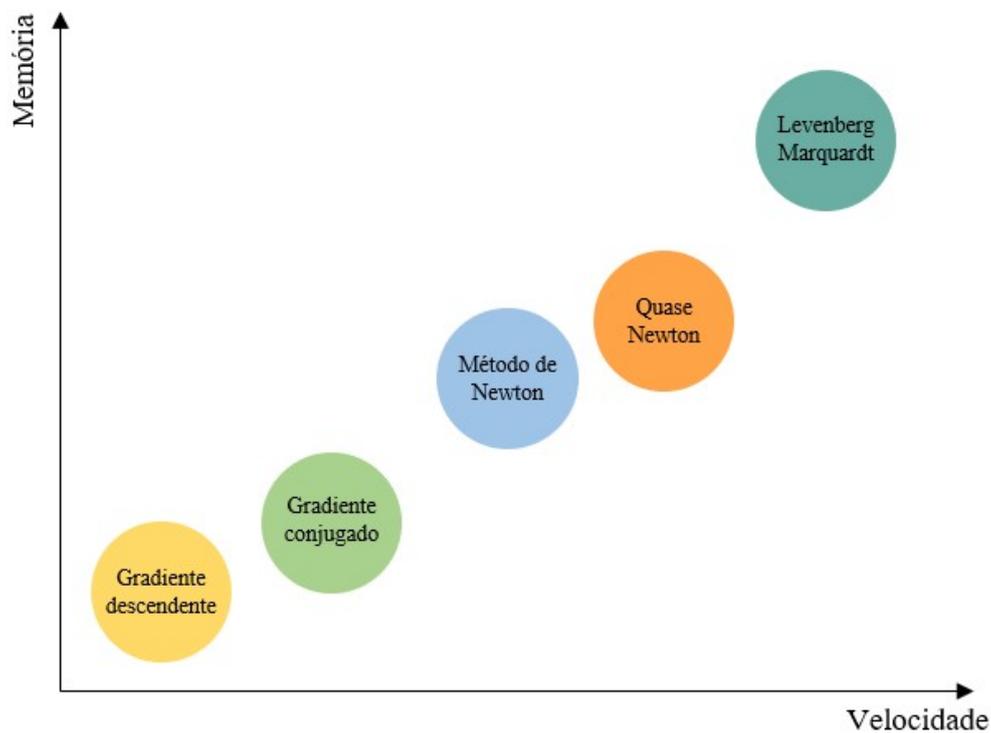


Figura 2.4 – Principais técnicas de otimização em Redes Neurais Artificiais

Fonte: Baseado em Haykin (2008)

O mais popular algoritmo de treinamento de uma MLP é o gradiente descendente implementado através da retropropagação do erro. Uma das razões para a frequente adoção dele é o baixo consumo de memória. Isso se deve ao fato de o Gradiente descendente exigir somente o armazenamento do vetor gradiente de tamanho  $n$ , em contrapartida aos métodos que demandam o armazenamento da matriz hessiana (tais como o Método de Newton, e o Quasi Newton) de tamanho  $n^2$  (HAYKIN, 2008). A retropropagação do erro (RUMELHART, HINTON, WILLIAMS, 1986), ou simplesmente *Backpropagation*, é um método supervisionado de treinamento de MLPs, baseado na regra de aprendizado de correção do erro (PRINCIPE, EULIANO, LEFEBVRE, 1999). A regra consiste na execução iterativa de duas etapas sobre as camadas da rede, primeiro um passo de processamento direto e, posteriormente, um de processamento reverso.

No primeiro passo do algoritmo um padrão é apresentado aos neurônios de entrada e o efeito é propagado camada por camada em direção à camada de saída. Uma vez que as saídas individuais dos neurônios de determinada camada, com exceção daqueles da camada de saída, estão conectadas às entradas da camada seguinte (Figura 2.3), devemos calcular a saída de todos os neurônios da camada atual antes que possamos passar ao processamento da camada seguinte, para assim dispormos de todos os valores das entradas da próxima camada antes que iniciemos o processamento dela. As saídas

produzidas pelos neurônios da última camada devem ser comparadas aos valores esperados, e assim obtemos o erro através da aplicação da equação 2.3.

Com a obtenção do erro, buscamos então a minimização da média quadrática do erro, denominada Erro Quadrático Médio (EQM), que leva em consideração o erro calculado a partir de cada neurônio da camada de saída. Isso resulta em uma nova função  $E(t)$ , que podemos vincular a cada iteração  $t$ :

$$E(t) = \frac{1}{N} \sum_{i=1}^N (e_i(t))^2 = \frac{1}{N} \sum_{i=1}^N (d_j - y_j(t))^2 \quad (2.7)$$

Com isso, passamos ao passo reverso do algoritmo, em que a regra de atualização dos pesos e bias é dada por:

$$w_{ij}^m(t+1) = w_{ij}^m(t) - \eta \frac{\partial E(t)}{\partial w_{ij}^m} \quad (2.8)$$

e

$$b_{ij}^m(t+1) = b_{ij}^m(t) - \eta \frac{\partial E(t)}{\partial b_{ij}^m} \quad (2.9)$$

O cálculo das derivadas da função  $E(t)$  para cada camada da rede é obtida através da aplicação da regra da cadeia, seguindo o fluxo do passo reverso, ou seja, partindo da camada de saída até a entrada. O valor de  $\eta$  é definido pelo usuário uma vez que esse consiste no hiperparâmetro taxa de aprendizado (Tabela 2.1). Os índices  $i$ ,  $j$  e  $m$  identificam o neurônio na rede, sendo o  $j$ -ésimo peso, do  $i$ -ésimo neurônio, da  $m$ -ésima camada. A aplicação da regra da cadeia se dá através de um processo de retropropagação do erro, termo esse reaproveitado para nomear o algoritmo.

No passo reverso o erro obtido na etapa anterior deve ser propagado de maneira retroativa na rede, ou seja, da última camada em direção à entrada. A seguir os pesos sinápticos são ajustados de forma a corrigir a resposta da rede em direção ao padrão esperado. A mudança nos pesos é procedida seguindo a regra de generalização do delta (HAYKIN, 2008) para minimização do erro.

### 2.3 Support Vector Machine (SVM) e Support Vector Regression (SVR): alguns apontamentos

A técnica de aprendizado de máquinas denominada Máquina de Vetor Suporte, ou do Inglês *Support Vector Machine* (SVM), foi apresentada por Boser, Guyon e Vapnik (1992), muito embora os conceitos que a sustentem não sejam novos, datando de 1960 (inclusive trabalhos anteriores de Vapnik e Chervonenkis sobre teoria de aprendizado estatístico) (HAN; KAMBER; PEI, 2011). Inicialmente concebida para classificação em duas classes (CORTES; VAPNIK, 1995), o SVM consiste em aplicar um mapeamento não linear sobre os dados de entrada de forma a elevá-los a um espaço dimensional mais alto. Nesse espaço, em que é possível estabelecer a separação entre as classes, são determinadas margens lineares (ou hiperplanos) para tal. As propriedades especiais do hiperplano asseguram a habilidade de generalização da técnica.

Na Figura 2.5 podemos observar dois exemplos de separação das classes. A classe a qual pertence cada observação ( $y = 1$  ou  $y = -1$ ) é dada pelos símbolos: quadrado e triângulo. Podemos traçar inúmeras linhas de forma a separar completamente os dois grupos de classes, todavia é necessário estabelecer um critério para que a quantidade ótima seja encontrada (GUENTHER; SCHONLAU, 2016).

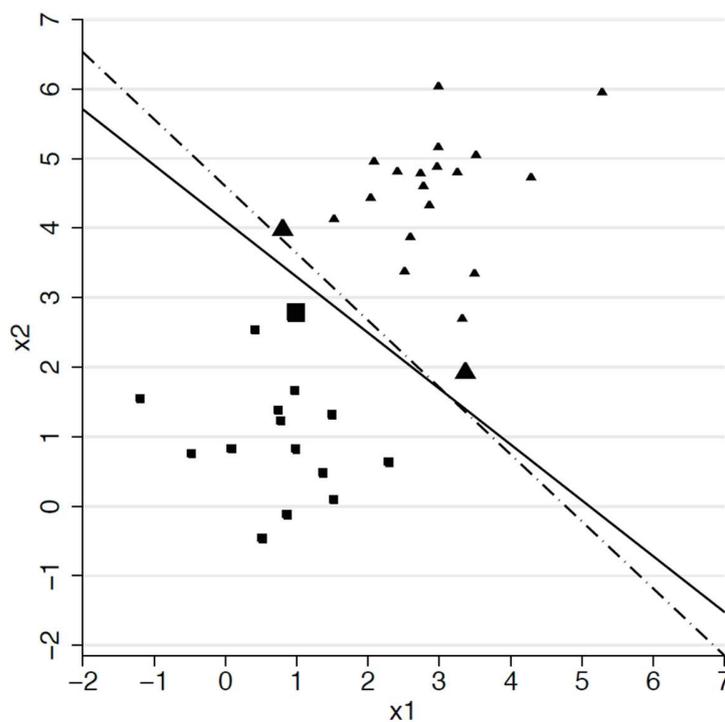


Figura 2.5 – Separação de duas classes.

Fonte: Guenther; Schonlau (2016)

A Figura 2.5 apresenta duas classes de pontos separáveis por margem máxima (linha sólida), e uma segunda fronteira de decisão (linha tracejada). Os triângulos e quadrados denotam as distintas classes (valores de  $y$ ). Os triângulos e quadrados maiores representam os vetores suporte.

Uma solução possível consiste na escolha da linha cuja margem (distância mínima entre a linha), ou fronteira de decisão, e o ponto mais próximo seja maximizada. Esta decisão leva ao seguinte problema de otimização (GUENTHER; SCHONLAU, 2016):

$$\begin{aligned} & \text{Max } M \\ & \beta_j, j = 1, \dots, p \end{aligned} \tag{2.10}$$

sujeito a

$$\sum_{j=1}^p \beta_j^2 = 1 \tag{2.11}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M \tag{2.12}$$

em que  $M$  é a margem,  $p$  é o número de  $x$  variáveis, e  $y_i \in \{-1, 1\}, i = 1, \dots, n$ . A equação 2.12 representa um hiperplano. Em duas dimensões, como na Figura 2.5, o hiperplano é uma linha. Os pontos acima do hiperplano devem satisfazer a condição:

$$(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) > 0 \tag{2.13}$$

os pontos abaixo do hiperplano devem satisfazer a condição:

$$(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) < 0 \tag{2.14}$$

Os pontos abaixo do hiperplano recebem o rótulo de classe  $-1$ , dessa forma:

$$y(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \tag{2.15}$$

que é sempre positivo, desde que as duas classes sejam separáveis.

Essa conveniência matemática é a razão pela qual são adotados  $\{-1, 1\}$  como rótulos, ao invés de outras possibilidades tais como  $\{0, 1\}$ , tipicamente utilizados no contexto da Regressão Logística. A restrição sobre  $\beta_j$  na Equação 2.11 não é necessária, mas traz a vantagem de permitir determinar a

distância de qualquer ponto  $i$  em relação ao hiperplano através da equação 2.15 quando  $\beta_j$  é dimensionada, desde que a restrição esteja presente (GUENTHER; SCHONLAU, 2016).

O problema de otimização das equações 2.10, 2.11 e 2.12 pode ser solucionado através de um algoritmo que trate problemas quadráticos. A solução depende somente das observações presentes na margem. Isso reduz o esforço computacional para solução, pois somente uma fração dos dados é necessária para computar uma solução ótima. Esse conjunto de observações é denominado vetores suporte (GUENTHER; SCHONLAU, 2016). O exemplo da Figura 2.5 apresenta três vetores suporte.

Em grande parte das aplicações, as duas classes não são separáveis e o problema de otimização não tem solução. Isso pode ser tratado permitindo uma certa quantidade de erros de classificação:

$$\begin{aligned} & \text{Max } M \\ & \beta_{ij}, \epsilon_i, i = 1, \dots, n; j = 1, \dots, p \end{aligned} \quad (2.16)$$

sujeito a

$$\sum_{j=1}^p \beta_j^2 = 1 \quad (2.17)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \quad (2.18)$$

$$\epsilon_i \geq 0 \text{ e } \sum_{i=1}^n \epsilon_i \leq C \quad (2.19)$$

A introdução das variáveis de folga  $\epsilon_i$  permite observações presentes no lado oposto da margem, ou seja, classificadas incorretamente. Todavia, a soma dos erros não pode exceder o limite máximo atribuído à  $C$ .

Novamente a solução para o problema de otimização depende somente dos vetores suporte, sejam eles observações presentes em uma ou outra margem ou aquelas que violaram a margem. Na Figura 2.6 a linha sólida divide quase que inteiramente as classes dos triângulos e dos quadrados, exceto por aqueles destacados por círculos. Adicionalmente, a linha tracejada superior consiste na margem para os triângulos, enquanto que a linha tracejada inferior é a margem para os quadrados. As observações que violam as margens não obrigatoriamente estão classificadas incorretamente, a não ser que, como no caso das envoltas por círculos, estejam no lado oposto ao grupo da classe ao qual pertençam em relação à fronteira de decisão. Por fim, o quadrado incorretamente classificado é considerado um vetor suporte que se encontra além da margem inferior, porém a violação é tão pronunciada que ele se encontra além da margem superior (GUENTHER; SCHONLAU, 2016).

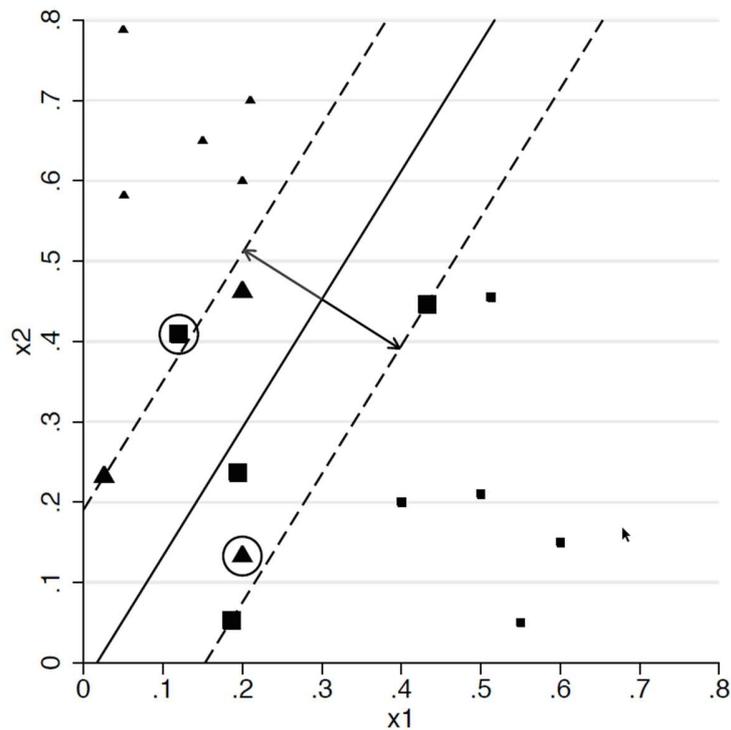


Figura 2.6 – Duas classes de observação não separáveis.

Fonte: Guenther; Schonlau (2016)

Na Figura 2.6, a linha sólida representa a fronteira de decisão; as linhas tracejadas representam as margens superior e inferior. Os triângulos e quadrados representam as duas classes (valores de  $y$ ). As formas maiores consistem nos vetores suporte. As observações classificadas incorretamente, ou seja, presentes no lado errado da fronteira de decisão, estão contornadas por círculos.

Valores mais altos atribuídos ao parâmetro  $C$  resultam em uma maior penalidade para classificações incorretas. Isso corresponde à definição de margens mais estreitas (BISHOP, 2006). Simetricamente, a consequência de um  $C$  mais alto é uma menor quantidade de vetores suporte. Isso ocorre porque quanto mais estreitas são as margens, um número menor de observações está sujeito à avaliação, seja por resultar em classificações incorretas ou simplesmente identificadas sobre as margens. Em último caso, um  $C$  elevado pode levar a um modelo super adaptado ao conjunto de treinamento, mas com performance baixa quando submetido ao conjunto de teste, que precisamente é o conceito de *overfitting*. Desta forma, o parâmetro  $C$  deve ser cuidadosamente calibrado através de uma técnica de validação cruzada.

Matematicamente, o processo de classificação em duas categorias de  $y$  é baseado em  $\text{sign}\{\hat{f}(x)\}$  que remete ao lado do hiperplano em que a observação se encontra, com:

$$\begin{aligned}\hat{f}(x) &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip} \\ &= \hat{\beta}_0 + \sum_{i \in S} \hat{\alpha}_i y_i \langle x, x_i \rangle\end{aligned}\quad (2.20)$$

Onde  $\hat{\alpha}_i > 0$  são coeficientes estimados, obtidos durante a otimização (BURGES, 1998),  $S$  é o conjunto de vetores suporte, e  $\langle \cdot, \cdot \rangle$  o produto (escalar) interno. O produto escalar é uma medida de similaridade entre as duas observações  $i \in \{1, \dots, n\}$  e  $i' \in \{1, \dots, n\}$ . Denominado de  $K$  ou também definido por kernel linear (BURGES, 1998), em que:

$$K(x_i, x_{i'}) = \langle x_i, x_{i'} \rangle = \sum_{j=1}^p x_{ij} x_{i'j} \quad (2.21)$$

Uma maneira de adicionar variáveis a um problema forma é através de uma função de variáveis originais, tais como variáveis polinomiais. As variáveis polinomiais aumentam o número de variáveis de uma regressão, entretanto o número de variáveis  $x$  subjacentes permanece o mesmo. Esse tipo de recurso também pode ser empregado no SVM. Dessa forma, a adição de variáveis não resulta em mudança no problema de otimização das equações 2.16 a 2.19.

Seja  $h(x_i) = \{h_1(x_i), h_2(x_i), \dots, h_{p'}(x_i)\}$  para  $i = 1, \dots, n$ . E tomemos como exemplo a inserção de um termo quadrático para  $x_1$ , de observação  $i$ , através da seguinte forma:  $h(x_i) = (x_{i1}, x_{i2}, \dots, x_{ip}, x_{i1}^2)$ . Dessa a forma a equação 2.20 poderia ser reescrita:

$$\begin{aligned}\hat{f}(x) &= \hat{\beta}_0 + \hat{\beta}_1 h(x_{i1}) + \dots + \hat{\beta}_p h(x_{ip}) \\ &= \beta_0 + \sum_{i \in S} \hat{\alpha}_i y_i \langle h(x), h(x_i) \rangle\end{aligned}\quad (2.22)$$

A solução para o problema de otimização da equação 2.22 depende somente do produto (escalar) interno das observações, e não das observações de fato.

A utilização do kernel linear na técnica SVM é especificada pela equação 2.21. Entretanto, grande parte dos problemas não são linearmente separáveis, dessa forma o uso de um kernel não linear confere mais flexibilidade uma vez que permite uma separação das classes desta outra gama de problemas. Os kernels não lineares mapeiam as variáveis  $p$  e  $x$  em um espaço de dimensão mais elevado. O mapeamento ocorre de maneira implícita; as variáveis não são de fato calculadas no espaço de mais alta dimensão. Ainda assim, uma quantidade constante de observações apresenta-se de maneira mais dispersa quando transportada para o espaço de mais alta dimensão, tornando mais fácil a separação linear. Isso melhora o desempenho do algoritmo de SVM e, como consequência, torna a fronteira de decisão no espaço original mais adaptável (HASTIE, TIBSHIRANI, e FRIEDMAN, 2009).

Um dos kernels não lineares mais utilizados, e também adotado por esse trabalho, é o kernel radial (ou gaussiano), também denominado *Radial Basis Function* (RBF):

$$K(x_i, x_{i'}) = \exp \left\{ -\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\} \quad (2.23)$$

em que  $\gamma > 0$  consiste no hiperparâmetro denominado *gamma*.

Quando uma observação de teste se encontra distante de uma observação de treinamento, o expoente torna-se fortemente negativo, e  $K(x_i, x_{i'})$  tende a zero. Nesse caso as observações de treinamento que se encontrarem longe de uma observação de teste, tem efeito reduzido sobre a classificação desta última. O hiperparâmetro  $\gamma$  regula quão perto uma observação deve estar para que sua influência contribua de maneira maior ou menor na decisão da classificação.

Além do RBF existem outras opções de kernels não lineares para utilização junto ao SVM. Uma dessas possibilidades é o kernel polinomial. A utilização desse kernel traz o hiperparâmetro  $d$ , a ser definido, que consiste no grau do polinômio que deve modelar a fronteira de decisão. Quanto maior for o grau do polinômio resultante de um valor elevado atribuído ao hiperparâmetro  $d$ , maior será a chance da criação de uma fronteira de decisão super-adaptada aos dados. A forma geral desse kernel é dada por:

$$K(x_i, x_{i'}) = \left( \beta_0 + \gamma \sum_{j=1}^p x_{ij} x_{i'j} \right)^d \quad (2.24)$$

em que  $\gamma > 0$  e  $\beta_0$  também são hiperparâmetros suportados por esse kernel. A semântica do hiperparâmetro  $\gamma$  é a mesma descrita no kernel RBF, enquanto que  $\beta_0$  influencia na parcialidade da métrica de similaridade de todas as amostras.

Por fim, menos utilizado do que os anteriores, em parte devido à algumas indesejáveis propriedades teóricas (LIN, LIN, 2003), temos o kernel sigmoide dado por:

$$K(x_i, x_{i'}) = \tanh\left(\beta_0 + \gamma \sum_{j=1}^p x_{ij}x_{i'j}\right) \quad (2.25)$$

em que os hiperparâmetros  $\gamma > 0$  e  $\beta_0$  são utilizados da mesma forma que no kernel polinomial.

### 2.3.1 Margem Suave

Consideremos o problema da classificação binária de  $m$  objetos que pertencem a dois conjuntos  $I_1$  e  $I_2$ . Cada objeto  $i$  ( $i = 1, \dots, m$ ) tem  $n$  colunas de caracterização (*features*) os quais estão armazenadas na  $i$ -th linha de uma matriz  $X$ ,  $m \times n$ . Para cada objeto  $i$ , definimos o rótulo de  $y_i$  da seguinte forma:  $y_i = 1$  se  $i \in I_1$ , ou  $y_i = -1$  se  $i \in I_2$  (LIN, LIN, 2003).

Através da técnica SVM é possível construir um hiperplano  $w^T x + b = 0$  que divide o conjunto de dados de forma a classificar os objetos  $x_i$  corretamente, ao tempo em que a margem de separação entre as duas classes  $\frac{2}{\|w\|}$  é maximizada (Figura 2.7).

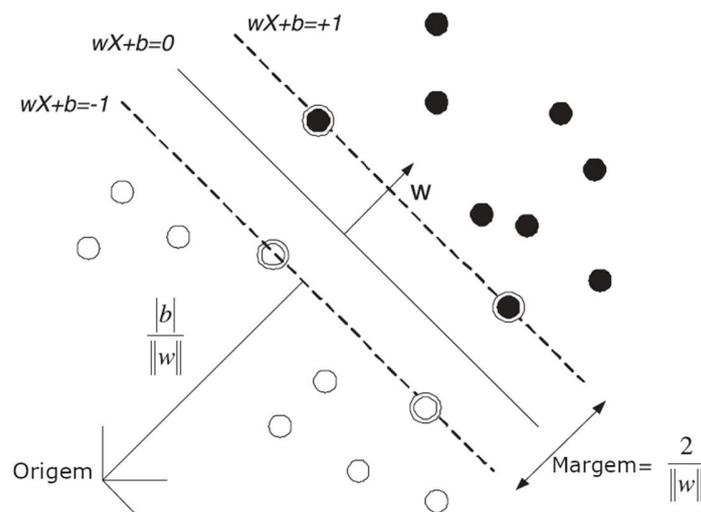


Figura 2.7 – Interpretação geométrica da margem suave para SVM.

Fonte: Maldonado, Merigó, Mirandas (2018)

A Figura 2.7 apresenta a representação geométrica do conceito de margem suave utilizado pela técnica SVM. Os vetores suporte encontram-se envoltos por circunferências. A maximização da margem é, teoricamente, motivada pelos limites de erro das propriedades de generalização da técnica. Desta forma, a probabilidade do erro de classificação de um objeto é dada por uma função que decresce em função do valor da margem (VAPNIK, 1998). O problema de programação quadrática abaixo é resolvido pelo SVM:

$$\text{Min}_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (2.26)$$

sujeito a

$$y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, \dots, m, \quad (2.27)$$

$$\xi_i \geq 0, \quad i = 1, \dots, m, \quad (2.28)$$

para uma dada amostra  $i$  do conjunto de treinamento,  $\xi_i$  representa quão distante aquela amostra se encontra no lado errado do hiperplano canônico (VAPNIK, 1998).

Para construirmos uma fronteira de decisão não linear através do SVM é preciso mapear as observações de treinamento em um espaço de dimensão superior, através de uma função kernel  $K(x_i, x_s) = \varphi(x_i) \cdot \varphi(x_s)$ , onde  $\varphi(\cdot)$  é a função de projeção. A partir da teoria da dualidade (BURGES, 1998), podemos fazer uso do kernel sem a necessidade de definir  $\varphi(\cdot)$ . Na formulação dual das equações 2.26 a 2.28, os exemplos de treinamento aparecem como produtos escalares, permitindo o uso da função kernel conforme citada anteriormente. O SVM com margem suave e utilização de função kernel, é dado por (SCHÖLKOPF, SMOLA, 2002):

$$\text{Max}_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,s=1}^m \alpha_i \alpha_s y_i y_s K(x_i, x_s) \quad (2.29)$$

sujeito a

$$\sum_{i=1}^m \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, m, \quad (2.30)$$

onde  $\alpha$  são os multiplicadores de Lagrange, relativos às restrições das equações 2.27 e 2.28.

### 2.3.2 SVR – A Versão para Regressão do SVM

A técnica Máquinas de Vetor de Suporte para Regressão foi apresentada por Smola e Schölkopf (1998), e já em sua forma original faz uso dos conceitos de SVM, tais como margem suave, para previsão de valores contínuos e séries históricas, e apresentado novamente em Smola e Schölkopf (2004). As principais vantagens do SVR sobre as técnicas contemporâneas de aprendizado de máquina consistem em: alto índice de acerto, mesmo em face a um número limitado de amostras, superação do problema do mínimo local, devido ao rigoroso uso da teoria da otimização da programação quadrática (TANG et al., 2020), e o uso do kernel para mapeamento de problemas não lineares para dimensão mais baixa, tal qual o SVM (LO; WANG, 2012).

No que diz respeito à solução ao problema da programação quadrática, necessária para modelagem da função que decresce em função do valor da margem, para se chegar à probabilidade do erro da regressão para um determinado valor, temos que, dado um vetor  $x_i \in \mathbb{R}^p, i = 1, \dots, n$  e um vetor  $y \in \mathbb{R}^n$ :

$$\text{Min}_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i + \xi_i^* \quad (2.31)$$

sujeito a

$$y_i - w^T x_i + b \leq \varepsilon + \xi_i, \quad (2.32)$$

$$w^T x_i + b - y_i \leq \varepsilon + \xi_i^*, \quad (2.33)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, m, \quad (2.34)$$

Nas equações 2.32 e 2.33 a penalização cujas amostras estão ao menos  $\varepsilon$  distantes do valor real. Essas amostras penalizam a função objetiva por  $\xi_i$  ou  $\xi_i^*$ , dependendo se as respectivas previsões se encontram acima ou abaixo de  $\varepsilon$ .

O problema da dualidade se dá por:

$$\text{Min}_{\alpha,\alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon e^T (\alpha + \alpha^*) - y^T (\alpha - \alpha^*) \quad (2.35)$$

sujeito a

$$e^T (\alpha - \alpha^*) = 0 \quad (2.36)$$

$$0 < \alpha_i, \alpha_i^* \leq C, i = 1, \dots, m \quad (2.37)$$

Onde  $e$  é o vetor de todos os uns,  $Q$  é uma matriz  $m \times m$  semidefinida positiva,  $Q \equiv K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  é o kernel. Neste ponto os vetores de treinamento estão implicitamente mapeados para um espaço dimensional mais elevado, através da função  $\phi$ .

Assim, a predição se dá por:

$$\sum_{i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (2.38)$$

### 3 METODOLOGIA

A descrição da metodologia foi dividida em três subseções, na quais: (1) descrição dos dados, (2) tratamento dos dados e (3) aplicação da técnica e refinamento.

#### 3.1 Descrição dos dados

O conjunto de dados (Apêndice 1) utilizado para desenvolvimento dos modelos de previsão de preço de fretes agrícolas contam com informações de rotas rodoviárias (origem/destino) para soja, milho e fertilizante, e um conjuntos de dados organizados em diferentes grupos em função de sua natureza (Figura 3.1) ao longo do período de 2012-2019.

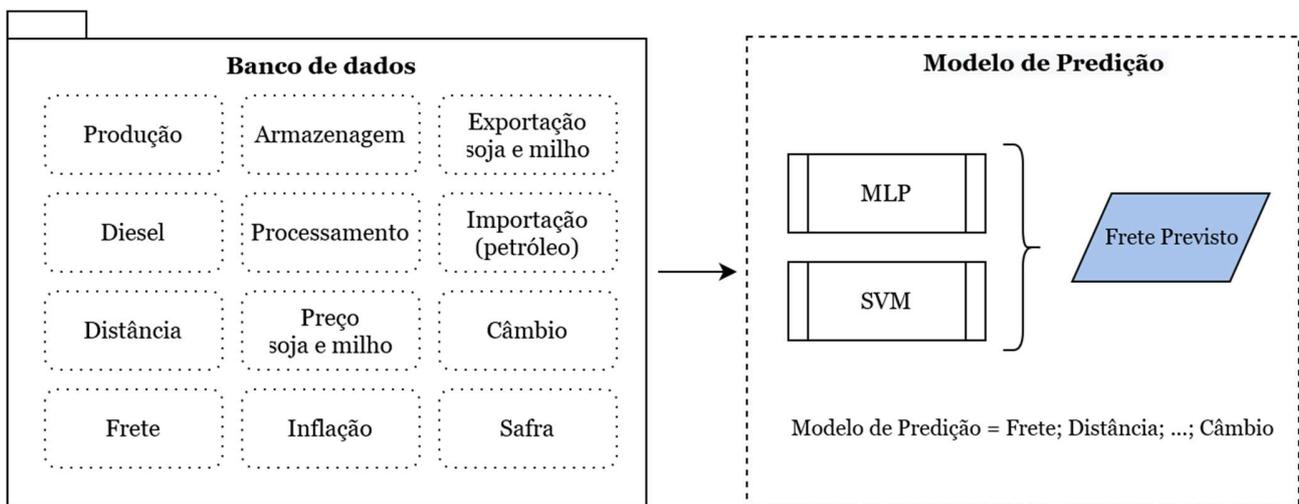


Figura 3.1 – Esquema do modelo de Predição de Fretes Rodoviários de Grãos

O grupo Frete compreende as informações mais essenciais ao modelo (Apêndice 1). Neste consta, além do valor  $Y$  que queremos prever (frete), as distâncias (km) entre as rotas praticadas (origem/destino) para os produtos soja, milho e fertilizante. A descrição dos grupos varia de acordo com o recorte geográfico, recorte temporal e unidade de medida (Apêndice 1).

## 3.2 Tratamento dos dados

Cada um dos conjuntos descritos na seção 3.1 estava dividido entre um ou mais arquivos. Em função disso, foi necessária uma fase de pré-processamento de dados para que as múltiplas fontes pudessem ser integradas em uma única estrutura, o que facilitou também a limpeza e identificação de *outliers*.

O primeiro passo do processo de integração foi a compilação das diferentes bases de dados em um formato detalhado único, além promover a correção dos dados, tais como: valores faltantes, formatados incorretamente (e.g.: numerais como texto), valores duplicados e/ou ambíguos. Tais cenários foram recorrentes e tratados da seguinte forma:

Para valores faltantes, o tratamento seguiu duas estratégias distintas, dependendo do conjunto em que foram encontrados, da seguinte forma: a. se identificados no conjunto de dados de Frete (Grupo I, Apêndice 1), a linha inteira foi removida pois esse conjunto detém dados essenciais, não permitindo a ausência da informação; b. se identificados nos demais conjuntos (Grupo II ao Grupo XI, Apêndice 1), e sabendo que todos os valores são numéricos, o campo vazio foi preenchido com o valor -1, sinalizando a ausência da informação.

A limpeza dos dados, necessária para remoção de *outliers* e de ruídos (ou valores incorretos), foi executada de maneira semiautomatizada. Foram gerados histogramas, um para cada coluna do conjunto de dados. No caso das colunas de tipo numérico, essas foram discretizadas em intervalos por intervalos, e então gerado o respectivo histograma. O histograma então foi submetido à inspeção manual, e o valor foi corrigido sempre que a informação estivesse disponível, caso contrário, a linha era inteiramente removida. No histograma gerado para a coluna preço do frete (R\$/t) (Figura 3. 2), o processo de inspeção concentrou-se principalmente nos intervalos mais elevados (destaque em vermelho, Figura 3.2), em busca de valores superestimados, possivelmente presentes em função de erro ocorrido durante a coleta dos dados. Sempre que a rota não justificasse o preço elevado foi considerado como erro, e tratado como tal.

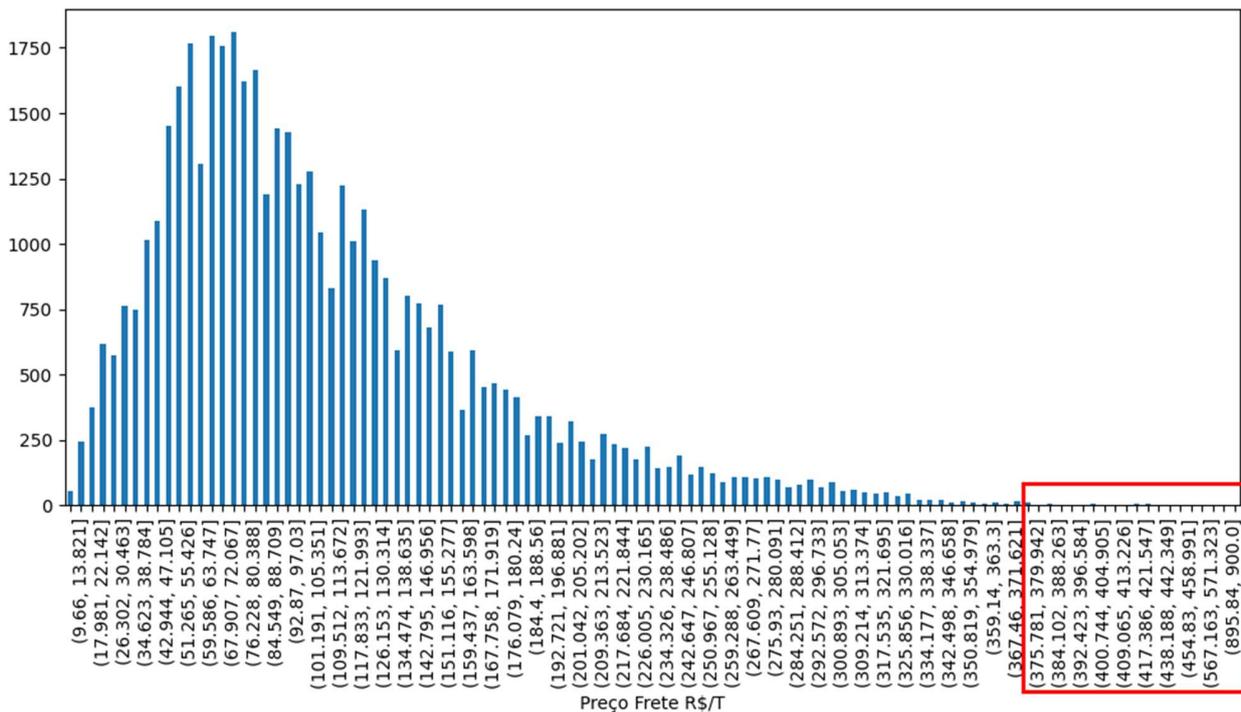


Figura 3.2 – Número de rotas por intervalo de preço de frete.

No último passo do pré-processamento dos dados procedemos com a integração das múltiplas fontes. Cada um dos grupos (Apêndice 1) foi recebida na forma de um arquivo. O Grupo I através das colunas: *commodity*, município de origem, UF origem, município de destino, UF destino, distância (em quilômetros), mês, ano, além do preço do frete (R\$/t), que é o valor que se deseja prever, ou o y dos dados, detém as informações que desenham a rota em si. Os demais dados (Grupo II à XI, Apêndice 1) continham as informações utilizadas para incrementar o conhecimento de cada uma das rotas presentes no Grupo I. Por isso algumas das características mais gerais do banco de dados são provenientes daí, como por exemplo o número total de linhas (ou rotas de frete) do banco de dados, o qual coincide com o número de registros do Grupo I, 46.185 registros.

Cada arquivo tratado nessa etapa foi processado por um *script* individual, desenvolvido na linguagem Python, e capaz de compreender as particularidades de cada conjunto de dados, e integrá-lo à matriz geral das informações de frete, cuja origem é o resultado do processamento do Grupo I. Apesar de diferentes implementações tenham sido necessárias para tratar as particulares organizações de cada arquivo, o algoritmo empregado para a integração foi em geral o mesmo, e seus passos são apresentados abaixo:

1. Adicionamos uma ou mais novas colunas vazias à matriz geral de rotas de frete, para que a informação da tabela do grupo II-XI possa ser adicionada.
2. Na tabela do grupo II-XI, identificamos manualmente as colunas cujos valores devam ser utilizados para busca de rotas na matriz geral
3. Submetemos cada linha da tabela corrente a uma busca por rotas na matriz geral, utilizando como critério de pesquisa os valores das colunas identificadas no passo anterior. O resultado deve ser um subconjunto maior ou igual a zero de rotas de frete originário da matriz geral
4. Fazendo uso do valor, da linha em análise, presente na coluna da tabela do grupo II-XI que queremos adicionar à matriz geral, preenchemos a coluna criada no passo 1, da matriz geral, nas linhas retornadas pela pesquisa do passo anterior
5. O processo é repetido para cada linha presente na tabela do grupo II-XI
6. Finalmente os valores não preenchidos (que não retornaram em momento algum pela pesquisa do passo 3.) na nova coluna recebem o valor -1, que o modelo deve perceber como ausência de informação.

O Diagrama UML de Atividade (Figura 3.3) ilustra como se deu o processo de integração do Preço do Diesel (Grupo V) à matriz geral. A geração desta última ocorre, primeiramente, a partir do processamento da Tabela do Grupo I, e então, cumulativamente, passa a receber as colunas adicionais a partir da análise das demais tabelas. No passo 1 é adicionada uma (ou mais) coluna vazia para receber o preço do diesel. No passo 2, que ocorre simultaneamente ao passo 1, são identificadas na Tabela do Grupo V as colunas cujos valores devem ser usados como critério de busca. No passo 3 a busca na matriz geral é executada fazendo uso do critério. O resultado é um subconjunto de linhas da matriz geral, que atendem ao critério. Cada linha desse subconjunto, na coluna preço do diesel, deve receber o valor do preço originário da Tabela do Grupo V (passo 4). Os passos 3 e 4 são repetidos para cada linha da Tabela do Grupo V (o passo 5 consiste no *loop*). Finalmente, através do passo 6, a coluna preço do diesel, da matriz geral, é inteiramente varrida, para que as células vazias (não retornadas por nenhuma busca do passo 3) recebem o valor -1, que simboliza para o modelo a ausência de informação.

Ao fim da aplicação de cada *script* de integração referente a cada uma das 10 tabelas, obtivemos uma matriz geral de dados composta por 49 colunas.

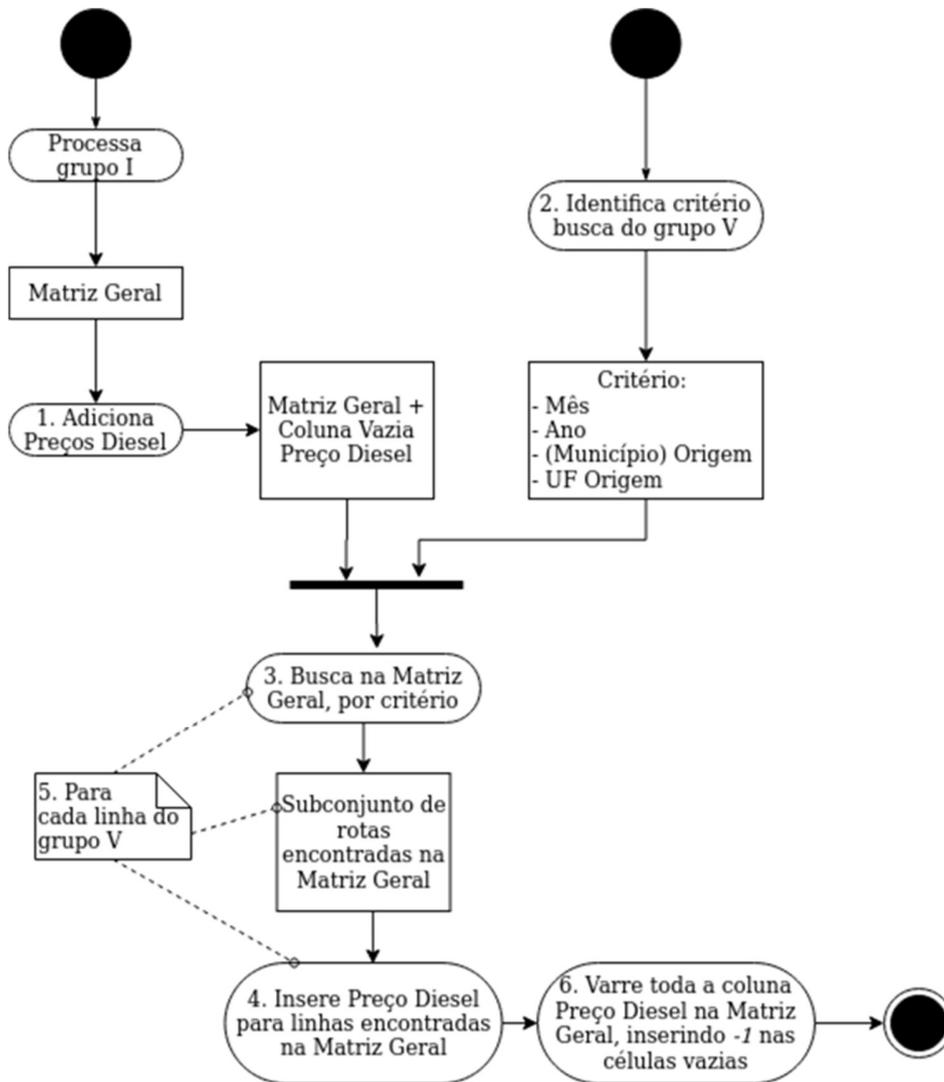


Figura 3.3 – Diagrama UML de Atividade do processo de integração

### 3.3 Aplicação da Técnica e Refinamento

A aplicação do procedimento de execução de cada uma das técnicas se deu de maneira independente, e foi mantido o mesmo para que a comparação entre resultados fosse possível. Naturalmente, ao longo de uma ou outra aplicação, pontos de melhoria foram identificados no processo (e.g. a utilização de *K-Fold* para Validação Cruzada), nesses casos a mudança foi adotada nos procedimentos da aplicação de ambas as técnicas, para que, novamente, os métodos se mantivessem os mesmos.

O processo de aplicação de cada técnica foi subdividido no conjunto de passos: (1) escolha das métricas, (2) validação cruzada, (3) análise de componentes principais, (4) remoção de fertilizantes, (5) particionamento da totalidade do conjunto de dados em dois subconjuntos disjuntos, o primeiro abrangendo as rotas até 600 quilômetros (inclusive), e o segundo detendo as rotas mais longas que 600 quilômetros (exclusive), (6) regularização (Figura 3.4).

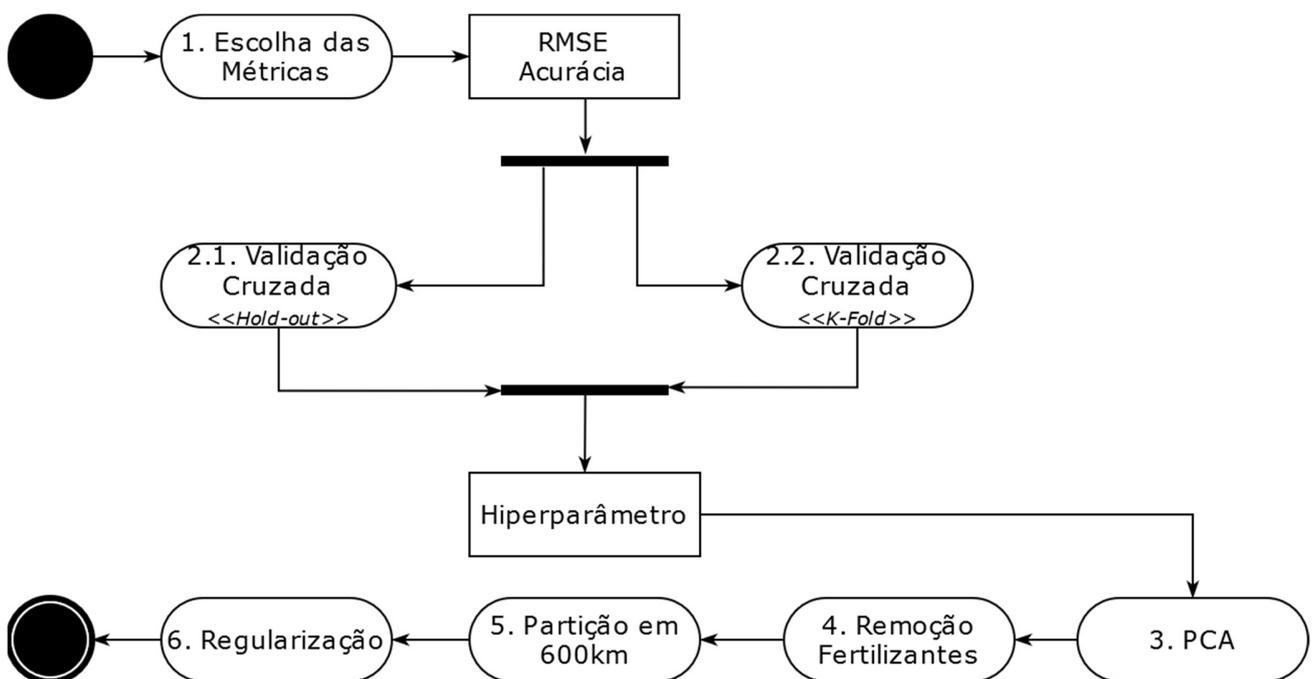


Figura 3.4 – Diagrama UML de Atividade da etapa de aplicação da técnica

Em linhas gerais o fluxo de execução se deu da seguinte forma: através do passo 1 (Escolha das Métricas) e dos passos 2.1 e 2.2 (Validação Cruzada) pudemos descobrir uma configuração eficiente de cada uma das técnicas selecionadas. A seguir os passos 3 (PCA) e 4 (Remoção de Fertilizantes) foram

executados de maneira interativa, repetidos e combinados, de forma a averiguar a necessidade da aplicação desses, em busca de valores aceitáveis para as métricas selecionadas no passo 1. Na etapa seguinte identificamos a necessidade de dividir o conjunto de dados em dois conjuntos, e assim o procedemos no passo 5. O ponto de quebra escolhido para geração dos dois conjuntos distintos foi a distância da rota percorrida entre a origem e o destino de 600 quilômetros.

Como identificado por Ferreira (2016), essa distância representa dois comportamentos diferenciados do preço do frete, um primeiro até 600 quilômetros, e o segundo a partir dessa distância. Sendo assim, para cada uma das subdivisões treinamos um modelo distinto, especializado naquela região dos dados, individualizada a partir desse momento. Por fim o passo 5 (Regularização) foi processado seguidas vezes, até que se fosse alcançado um resultado aceitável para as métricas de avaliação. A sistematização da execução de cada passo é detalhada nas subseções que se seguem (Figura 3.4).

### 3.3.1 Escolha das Métricas

Antes da implementação e aplicação da técnica foi necessário a escolha das métricas para avaliação do desempenho do modelo. A escolha foi pelo par de R-quadrado ( $R^2$ ) e raiz quadrada do erro quadrático médio (RMSE, do inglês *Root Mean Squared Error*), utilizadas em conjunto para avaliação de cada execução de cada modelo gerado. No período que precedeu a escolha da métrica, consideramos a utilização do erro absoluto médio (MAE, do inglês *Mean Absolute Error*). Todavia, pela necessidade de se atribuir penalizações mais elevadas aos preços de frete previstos mais distantes dos reais, optamos pelo uso do RMSE. Dito isso, e tendo em mente a conhecida sensibilidade do RMSE aos *outliers* (CHAI, DRAXLER, 2014), adotamos uma etapa de detecção e remoção de outliers, conforme descrita na seção 3.2, para que essa condição da métrica pudesse ser contornada já na fase de pré-processamento dos dados. Por fim, a escolha do  $R^2$  se deu pela necessidade de dispormos um índice que representasse o acerto do modelo em relação aos valores reais. Essa métrica assume o valor 1,0 para um modelo que sempre acerta, e 0 (zero) para um modelo que retorna um valor constante, independente da entrada. Além disso o  $R^2$  pode assumir valores negativos, pois o modelo pode apresentar performance inferior ao retorno da constante.

### 3.3.2 Validação Cruzada

A validação cruzada foi necessária para descoberta do valor atribuído aos respectivos hiperparâmetros de cada técnica, i.e., número de neurônios da camada escondida no caso da MLP; abertura do kernel no caso da SVR, além do hiperparâmetro de regularização em ambas as técnicas, que levassem ao menor erro e maior  $R^2$ . O método empregado com sucesso nesta etapa da pesquisa foi *Hold-out*. A respeito da divisão dos conjuntos de treinamento, teste e validação foi adotada a proporção 90%/10% para criação dos dois primeiros: treinamento e teste, e 70%/30%, aplicada sobre o maior conjunto resultante da divisão anterior (90%, que é o conjunto de treinamento), para criação dos conjuntos de treinamento e validação (Figura 3.5). O conjunto de teste (10% na primeira divisão) foi mantido fora e isolado das etapas de Validação Cruzada e Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) (Figura 3.4), tendo sido utilizado apenas a partir da etapa de Remoção de Fertilizantes (Figura 3.4). As execuções das técnicas foram parametrizadas com os valores ideais de hiperparâmetros encontrados pela técnica de validação cruzada.

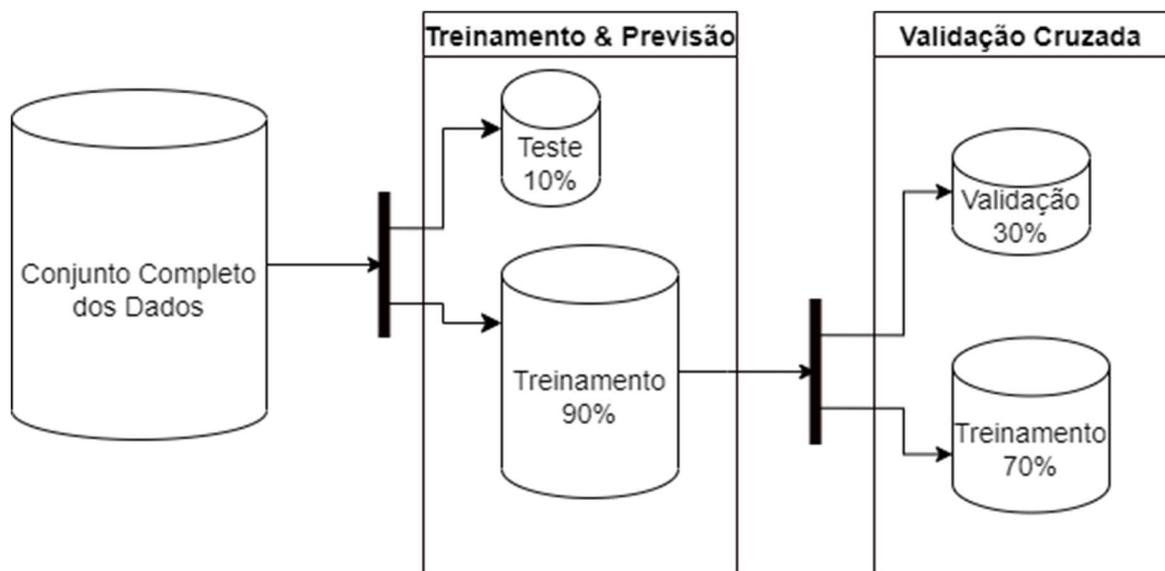


Figura 3.5 – Esquema de divisão do conjunto de dados em subconjuntos de treinamento, validação e teste

Alternativamente, a busca pelos hiperparâmetros também se deu através do método de Validação Cruzada *K-Fold*. Para isso, empregamos o valor 10 para o número de *folds*, posteriormente reduzido para 5. A utilização desse método apresentou resultados muito semelhantes ao *Hold-out*, mas fazendo uso de

um tempo relativamente mais longo. Enquanto que o *Hold-out* levou, no máximo, 3 horas para finalizar a busca pela quantidade de neurônios da camada escondida, no uso da MLP, o *K-Fold* levou aproximadamente 3 dias, inviabilizando dessa forma o seu uso pela limitação do tempo.

Seguindo a mesma linha de Josse e Housson (2012), a Validação Cruzada também foi empregada na busca pelo melhor valor associado ao hiperparâmetro variância, entrada do algoritmo de Análise de Componentes Principais. Com a diferença que no trabalho em questão, os autores obtiveram bons resultados fazendo uso do método para encontrar o número de componentes informado ao algoritmo, em lugar da variância.

### 3.3.3 Análise de Componentes Principais

Para ampliar as análises sobre o banco de dado foi aplicada a Análise de Componentes Principais (PCA) para investigar se o número de componentes, que totalizava 49, presentes no conjunto de dados poderia estar prejudicando a performance da técnica. Além disso, era preciso verificar se havia um subconjunto de colunas que pudesse ser utilizado ao invés do conjunto completo, mas que ainda assim mantivesse a representatividade. De fato, Ganguli e Dunnmon (2017) puderam reduzir de 12 para apenas 3 componentes, sem perda de representatividade, na aplicação de um modelo especializado em prever preço de títulos. Ainda, eles obtiveram um ganho de 75% na velocidade de execução do modelo. A ideia aqui foi reproduzir um ganho nessa mesma direção.

### 3.3.4 Remoção de Fertilizantes

A remoção dos dados de frete de fertilizantes deu-se em primeiro lugar como uma tentativa de melhorar as métricas num período pré-normalização dos dados. Outro fator relevante foi a presença de valores incorretos e inconsistentes nesse subconjunto dos dados. Por fim, devido ao comportamento do preço do frete de fertilizantes, que apresenta uma característica distinta quando em comparação aos preços das *commodities* soja e milho especialmente no que tange origem e destino, poderia prejudicar a capacidade de previsão do modelo (SAAB; PAULA, 2008).

A retirada dos dados de fertilizantes, apesar de reduzir o volume de dados de 46.184 para 31.714, elevou as métricas, levando a  $R^2$  para 0,96, e o RMSE para 14,25, como resultado da MLP, e 0,96 de  $R^2$ , e 13,87 de RMSE, no caso da SVR.

### 3.3.5 Particionamento do Conjunto de Dados

A necessidade de se particionar o universo dos dados em dois conjuntos disjuntos surgiu a partir da constatação de um  $R^2$  mais baixo para as rotas de menor distância. De fato, antes da regularização o  $R^2$  para as rotas até 600 quilômetros, obtida através da técnica MLP, foi de 0,81, enquanto que para as rotas acima de 600 quilômetros foi de 0,94. Todavia são as rotas mais curtas são justamente as mais movimentadas em todo o conjunto das rotas de frete, sendo assim apresentam-se em maior quantidade, o que torna essencialmente importante que o modelo apresente um maior índice de acerto para essas observações.

Trabalhos anteriores já identificaram que a curva do preço do frete apresenta comportamento diferente para distâncias mais curtas, quando em comparação às distâncias mais longas. Enquanto Ferreira (2016) fez uso dos intervalos, até, e a partir de 600 quilômetros, Oliveira et al., (2021) adotaram o particionamento em 500 quilômetros. Por outro lado, Soares e Caixeta (1997) implementou a divisão em intervalos de 50 quilômetros e Gonçalves, et al., (2014) empregaram intervalos de 200 quilômetros. Para o nosso caso, a divisão se deu em dois intervalos tendo a distância de 600 quilômetros como marco fez mais sentido, uma vez que justamente esse valor marca a mudança na tendência do preço das rotas de frete observada em nossa massa de dados.

### 3.3.6 Regularização

A remoção dos dados de fertilizantes levou as métricas, que já estavam nessa direção, ainda mais para o campo do sobre-ajuste (*overfitting*). Dessa forma, decidimos pela adoção de uma técnica de regularização para preservar a capacidade de generalização do modelo. Foi empregada a penalização da norma  $L2$  do vetor de parâmetros. Essa técnica é conhecida pelo nome em Inglês *Ridge Regression*.

Para se chegar a um valor eficaz para o parâmetro da penalização adotamos uma abordagem interativa, através do qual o valor foi variado num intervalo de 0,001 a 1.000 para a técnica MLP, e 1

a 1.000 para a técnica SVR, e foram verificadas as métricas resultantes de cada execução até que o  $R^2$  alvo, fixado em 0,85, tivesse sido atingido.

## 4 RESULTADOS E DISCUSSÕES

Com objetivo de obter um maior entendimento das variáveis e suas possíveis relações com o preço do frete dos grãos brasileiros, foi realizada uma análise preliminar das variáveis do banco. Analisando o comportamento dos preços de frete dos grãos ao longo do tempo, bem como o volume de fretes (Figura 4.1), notamos que entre 2013 e 2015 os dados apresentam uma maior quantidade de outliers (observações discrepantes quando comparadas às demais) em relação aos anos posteriores de 2016 a 2019. Somado a isso, uma maior variabilidade dos dados nos anos subsequentes à 2014, mais especificamente nos anos de 2016 e 2017. Observamos também a volumetria total de fretes mensais calculados entre os anos anteriores a 2015, no comparativo com este e os demais. Indicando assim, uma proporção distinta na base de dados entre os anos.

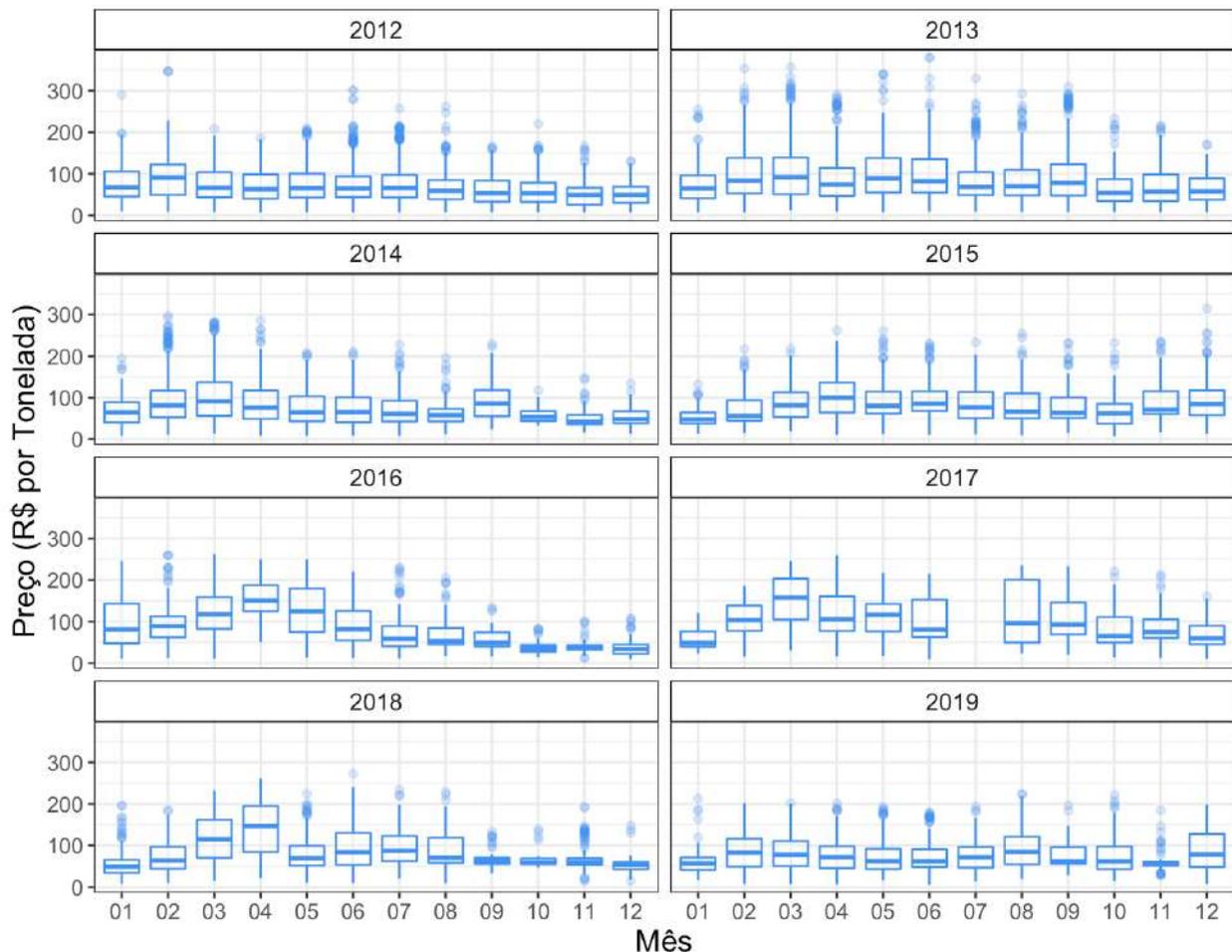
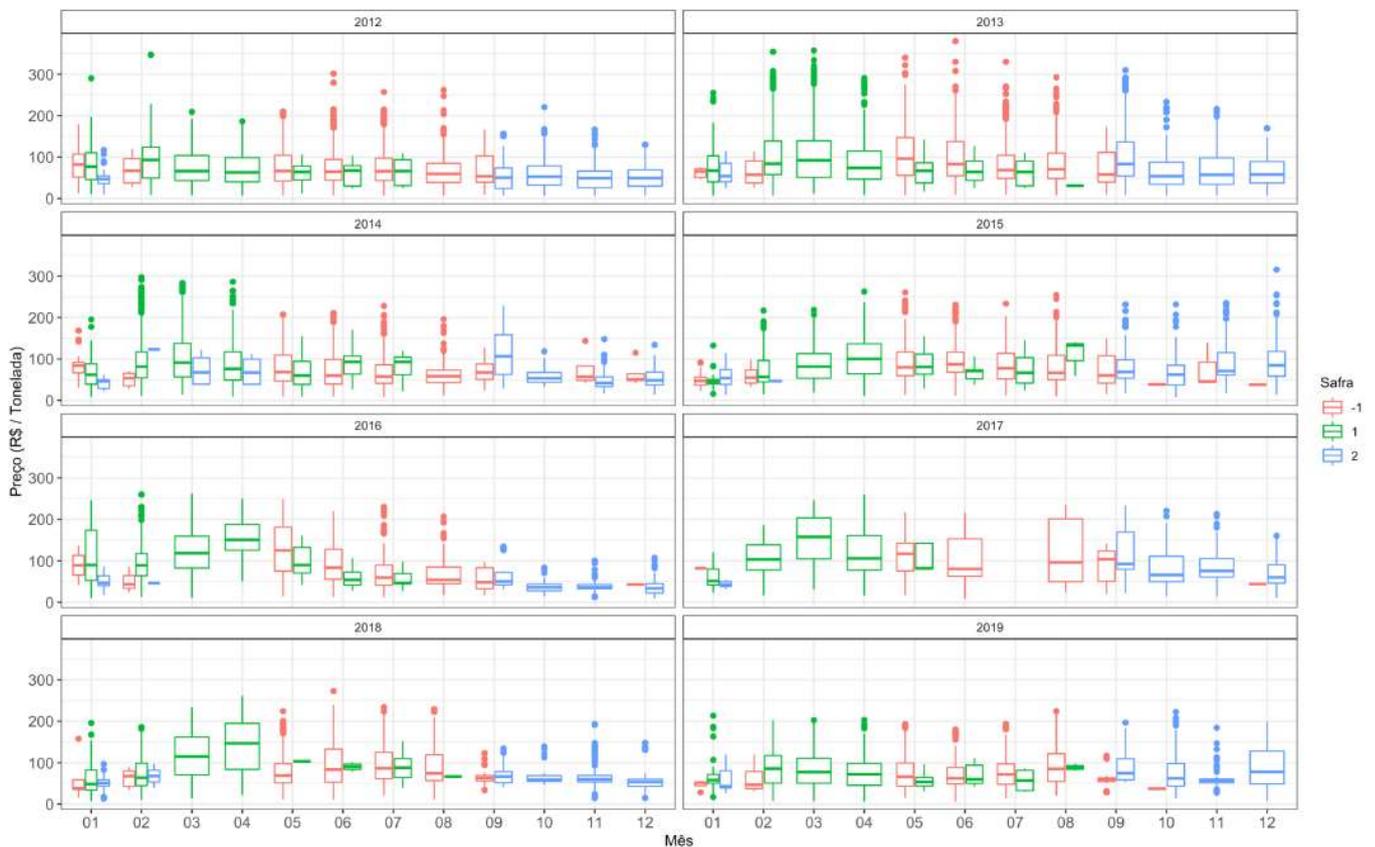


Figura 4.1 Box Plot do preço do frete da soja ao longo dos meses, 2012 - 2019

Os preços foram avaliados a partir do período da safra (Figura 4.2), com um comportamento estratificado entre as diferentes categorias, observamos uma maior ocorrência da entressafra, colheita e plantio, no meio, início e final do ano respectivamente. Identificando uma maior mediana no período de colheita. Desse modo, temos que os resultados provenientes dos testes indicam associação entre os pares de variáveis, corroborando em suma à relevância das variáveis explicativas para descrever a variável monetária.



Legenda: -1: entressafra, 1: colheita e 2: plantio

Figura 4.2 Box Plot do preço do frete da soja ao longo dos meses e período da safra, 2012 - 2019

#### 4.1 Resultados da Aplicação da MLP

Conforme mencionamos na seção 3.3 Aplicação da Técnica e Refinamento, a metodologia adotada para aplicação da técnica MLP foi subdividida em duas etapas. Na primeira etapa um único modelo foi criado, treinado, parametrizado e utilizado sobre ambas as partições dos dados (Figura 3.4), enquanto que na segunda foram dois os modelos, um para cada conjunto de dados. Em ambas as etapas

o processo de descoberta dos valores ideais para os hiperparâmetros, através da validação cruzada, e posterior previsão dos valores por cada modelo, foram executadas de maneira individualizada. Por fim, tanto o modelo único quanto os modelos individuais foram treinados fazendo uso de seus respectivos valores de hiperparâmetros descobertos no passo anterior e, a partir deles, foram obtidos os valores previstos para cada partição do conjunto de dados, totalizando quatro conjuntos de métricas: um para cada partição quando submetido ao modelo único, e um para cada partição obtidos a partir dos modelos individuais.

#### 4.1.1 MLP Única – Um Modelo para Duas Partições

O particionamento dos dados trouxe a necessidade de avaliação das métricas – RMSE e  $R^2$  – para cada um dos subconjuntos dos dados. No caso do modelo único, a rede foi treinada sobre o conjunto de rotas sem particionamento, ou seja, as rotas de qualquer distância presentes no conjunto, porém as métricas foram calculadas para cada partição individualmente. Tendo isso em mente, podemos perceber na

Figura 4.3 que a técnica de validação cruzada apresentou o número 91 como valor para o hiperparâmetro quantidade de neurônios na camada escondida. Isso ocorre porque foi justamente com aquele valor que o modelo MLP obteve o maior valor para o  $R^2$  (0,8345) e menor RMSE (9,8866), para as rotas de distância menor ou igual a 600 quilômetros.

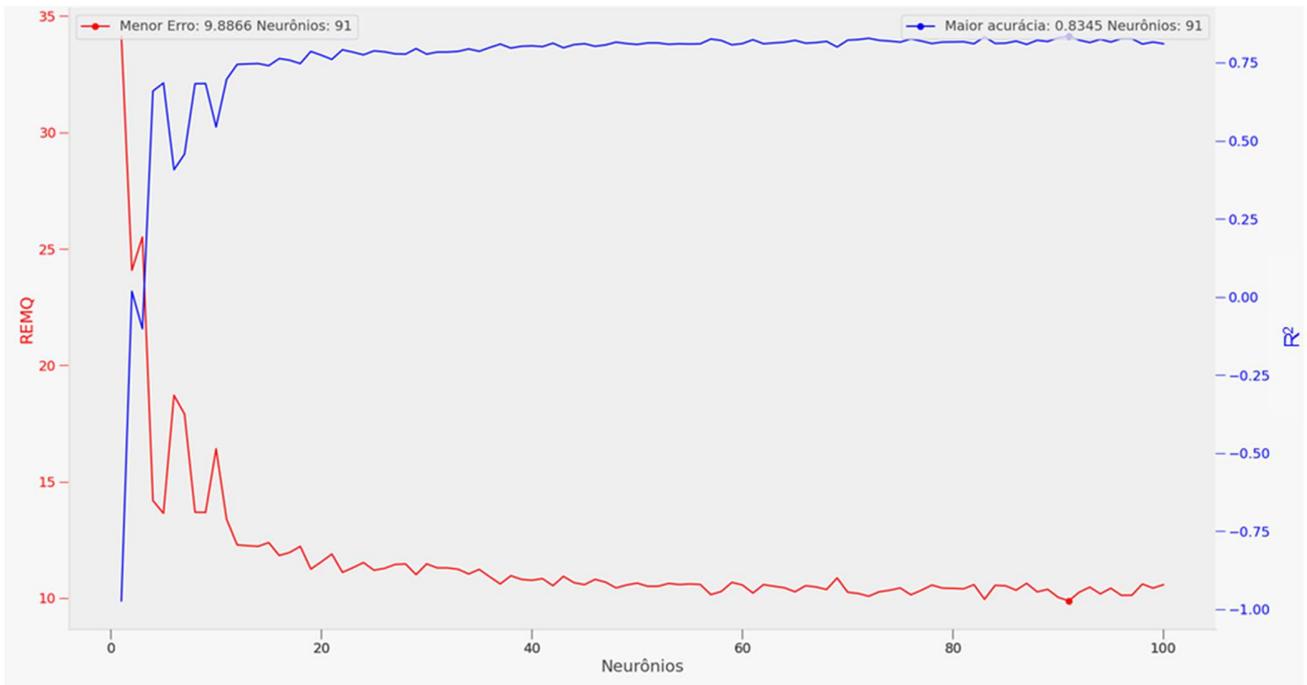


Figura 4.3 – RMSE e  $R^2$  para distância  $\leq 600$  km

O modelo configurado com os mesmos hiperparâmetros, retornou um  $R^2$  de 0,9296 e RMSE 16,9265, para as rotas de distância maiores que 600 quilômetros.

Tabela 4.1 – Resultados do *Hold-out* para busca do número de neurônios ideal

Otimização	Neurônios	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
MLP rotas curtas	91	0,8345	9,8866	0,9296	16,9265
MLP rotas longas	88	0,8214	10,2715	0,9342	16,3681

Por outro lado, quando as métricas a serem otimizadas são as da segunda partição (rotas de distâncias mais longas que 600 quilômetros), conforme podemos visualizar na Figura 4.4, a quantidade de neurônios ideal retornada pela técnica é 88, valor esse que implica em um  $R^2$  de 0,8214, e RMSE 10,2715, para a primeira partição; e 0,9342, e 16,3681, para a segunda, nas respectivas métrica. Dessa forma, notamos já na etapa de pesquisa dos valores para os hiperparâmetros, que, ao optarmos pela utilização de um modelo único, não foi possível obter um valor para o hiperparâmetro que resulte no melhor par de métricas para ambas as partições ao mesmo tempo. Isso pode ser constatado na Tabela 4.1 em que o maior  $R^2$  e menor erro para a primeira partição é obtida a partir de uma MLP com 91 neurônios,

enquanto que o melhor  $R^2$  e erro para a segunda partição é retornada de uma MLP diferente, configurada com 88 neurônios.

Essa característica manteve-se presente ao longo de todas as etapas de utilização de MLP modelo único. De qualquer maneira, a partir desse ponto todas as MLPs dessa seção foram configuradas com 91 neurônios na camada escondida, em função do retorno das melhores métricas com essa configuração, para a partição das rotas mais relevantes.

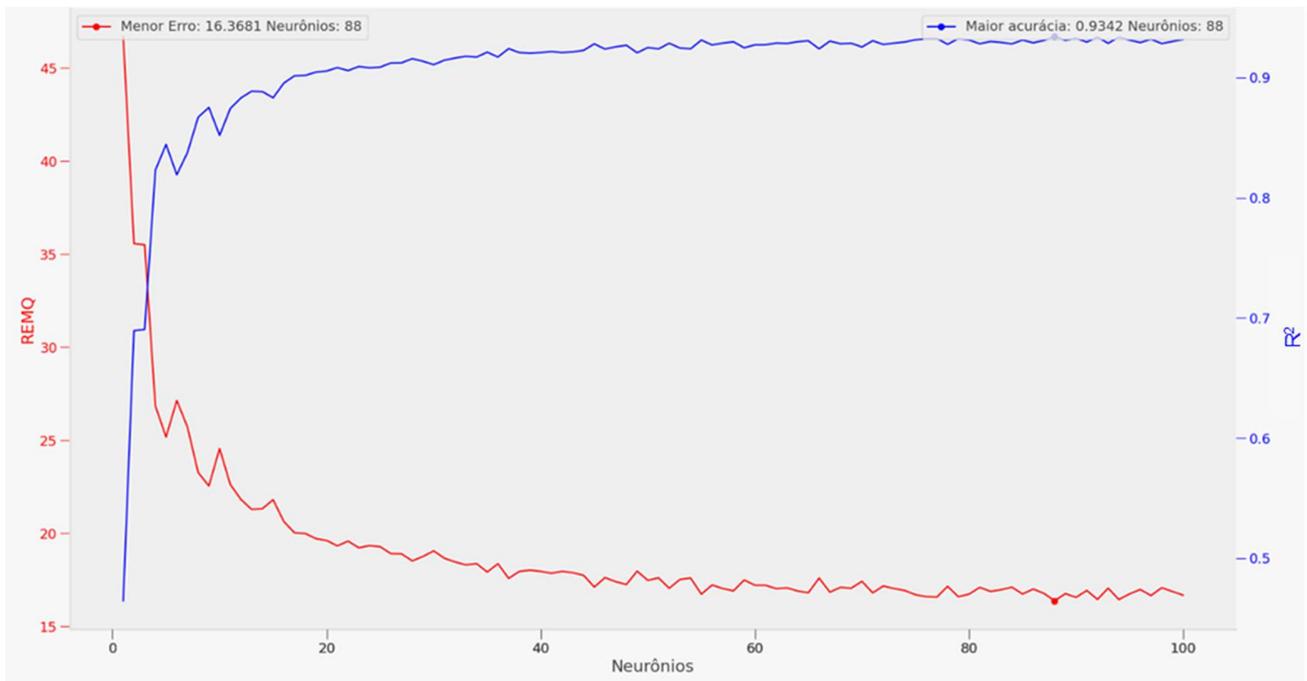


Figura 4.4 – RMSE e  $R^2$  para distância > 600 km

Após a identificação da quantidade ideal de neurônios na camada escondida, procedemos com a etapa de redução da dimensionalidade dos dados através da técnica PCA. No intuito de obter o valor ideal para a variância do método, empregamos novamente o *Hold-out* fazendo uso do conjunto de treinamento e do conjunto de validação. Duas foram as rodadas de execução da validação cruzada, a primeira tendo como objetivo a otimização das métricas sobre a partição das rotas curtas, e a segunda para otimizar as métricas da partição das rotas longas.

A primeira resultou em uma variância de 0,97,  $R^2$  de 0,8243 e RMSE de 10,1885 (Tabela 4.2). Foram selecionados 14 componentes: armazéns no município de destino, capacidade instalada indústria moagem no município de destino (t/dia), armazéns no município de origem, mercado internacional Chicago (US\$/dia), mercado internacional paridade (R\$/saca 60kg), distância (km), capacidade instalada indústria moagem no município de origem (t/dia), commodity, gasolina preço mínimo (R\$/l), valor da produção no ano (multiplicado por R\$ 1.000,00), rendimento médio (kg/ha), capacidade armazenagem municipal origem (t), destino, atacado (R\$/saca 60kg).

A segunda rodada de execução resultou em uma variância de 0,99,  $R^2$  0,9327, RMSE de 16,5544, e os componentes selecionados totalizaram 17: armazéns no município de destino, capacidade instalada indústria moagem no município de destino (t/dia), armazéns no município de origem, mercado internacional Chicago (US\$/dia), mercado internacional paridade (R\$/saca 60kg), distância (km), capacidade instalada indústria moagem no município de origem (t/dia), commodity, gasolina preço mínimo (R\$/l), valor da produção no ano (multiplicado por R\$ 1.000,00), rendimento médio (kg x ha), capacidade armazenagem municipal origem (t), destino, atacado (R\$/saca 60kg), capacidade armazenagem municipal no destino (t), mercado nacional (R\$/saca 60kg), mês. Ambos os resultados são apresentados na Tabela 4.2.

Tabela 4.2 – Resultados do *Hold-out* para busca da variância ideal

Otimização	Variância	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
MLP rotas curtas	0,97	0,8243	10,1885	N/A	N/A
MLP rotas longas	0,99	N/A	N/A	0,9327	16,5544

Apesar de uma diminuição em mais da metade do número de componentes originais, acompanhada de métricas muito próximas daquelas obtidas sem a seleção resultante do PCA, optamos pela manutenção de todas as colunas, por entendermos que a forma original dos dados, apesar de pouca diferença, ainda assim levou a melhores valores de  $R^2$  e RMSE, sem trazer uma perda de performance.

A etapa de calibragem do hiperparâmetro da penalização (L2) consiste no último uso da validação cruzada empregado em conjunto ao modelo único de MLP. Para tal, uma vez mais lançamos mão da técnica *Hold-out*. A primeira execução, através do qual buscamos a otimização do par de métricas para as rotas curtas, retornou o valor de 0,1 para a penalização (Figura 4.5). Esse valor, mesmo sendo a

penalização mínima resultou em um  $R^2$  de 0,8388, o que apesar de próxima, não atingiu o valor alvo (0,85) pré-definido para a métrica.

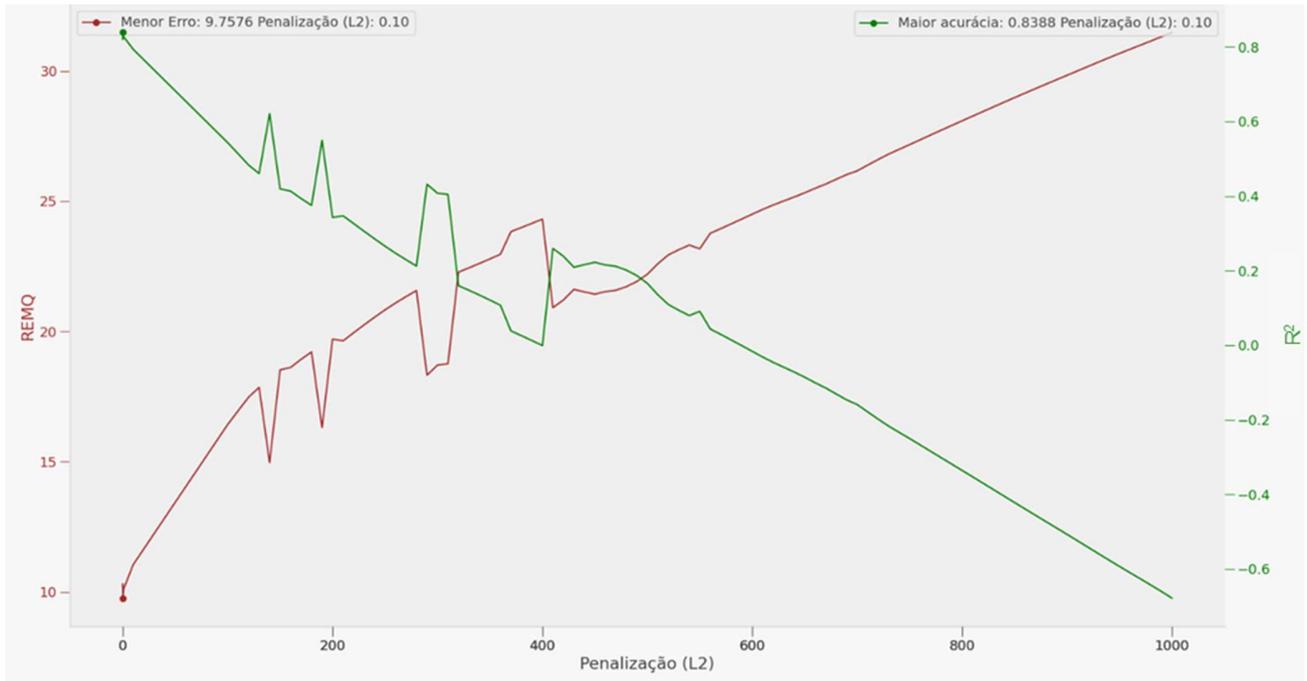


Figura 4.5 –  $RMSE$  e  $R^2$  para distância  $\leq 600$  km com penalização L2.

Subsequente a segunda execução, através do qual buscamos a otimização das métricas das rotas longas, retornou o resultado de 100,0 para a penalização (Figura 4.6), valor esse que aplicado à técnica de *Hold-out* resultou em um  $R^2$  de 0,8390, e RMSE de 25,6067.

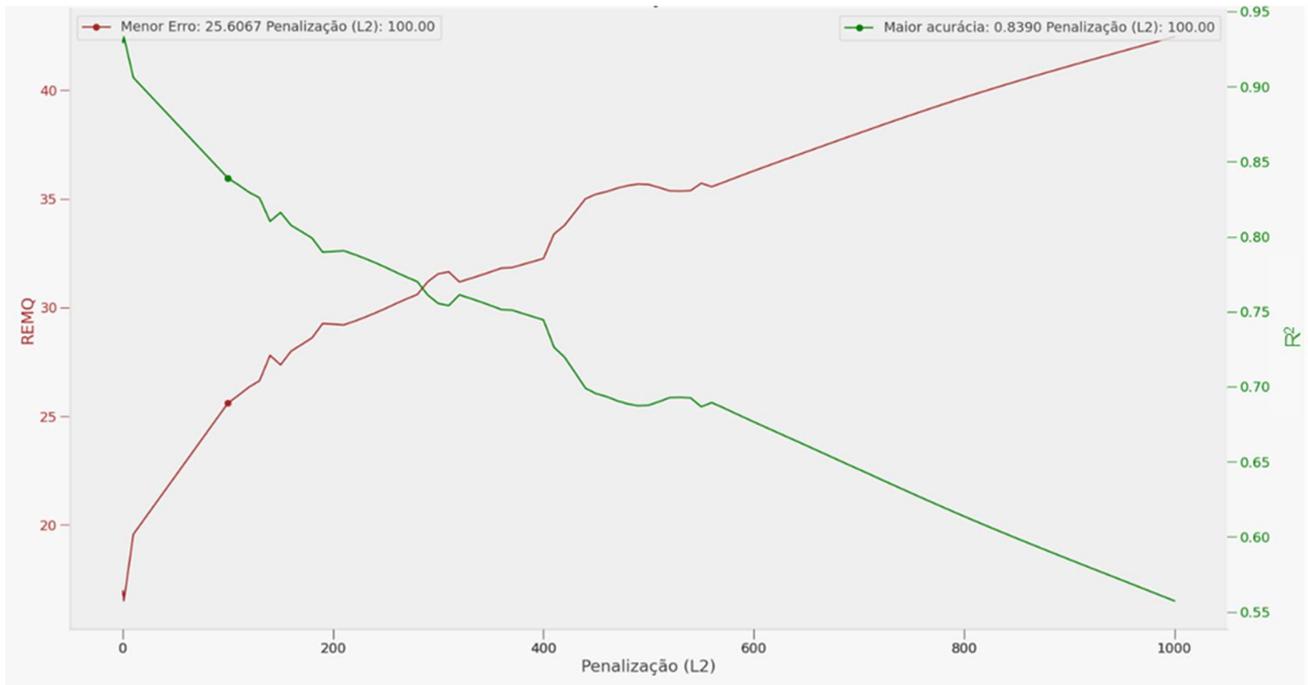


Figura 4.6 –  $RMSE$  e  $R^2$  para distância > 600 km com penalização L2.

Quando analisamos o comparativo das métricas na Tabela 4.3, observamos novamente o recorrente problema da utilização do modelo único. Quando reduzimos a penalidade buscando o resultado almejado para as rotas curtas, elevamos o  $R^2$  das rotas longas (0,9293), causando um potencial sobre-ajuste nos dados daquela partição. Por outro lado, ao elevarmos a penalidade para mover a métrica das rotas longas para o valor que definimos, causamos um sub-ajuste (*underfitting*, com  $R^2$  à 0,5434) às rotas curtas.

Tabela 4.3 – Resultados do *Hold-out* para busca do valor da penalização (L2) ideal.

Otimização	L2	$R^2_{\leq 600}$	$RMSE_{\leq 600}$	$R^2_{> 600}$	$RMSE_{> 600}$
MLP rotas curtas	0,1	0,8388	9,7576	0,9293	16,9726
MLP rotas longas	100,0	0,5434	16,4222	0,8390	25,6067

A variável distância é a que mais interfere no comportamento do valor do frete, já observado por Cangussu et al. (2013), Martins (2008), Corrêa Junior e Caixeta Filho (2003) e Mc Cann (2001). Isso

pode ser observado também através da análise da matriz de correlação de Pearson (Figura 4.7), no recorte da matriz a variável Distância (km) apresenta uma relação crescente mais forte (0,90) para com o Preço do Frete (R\$/t), dentre todas as variáveis do banco de dados.

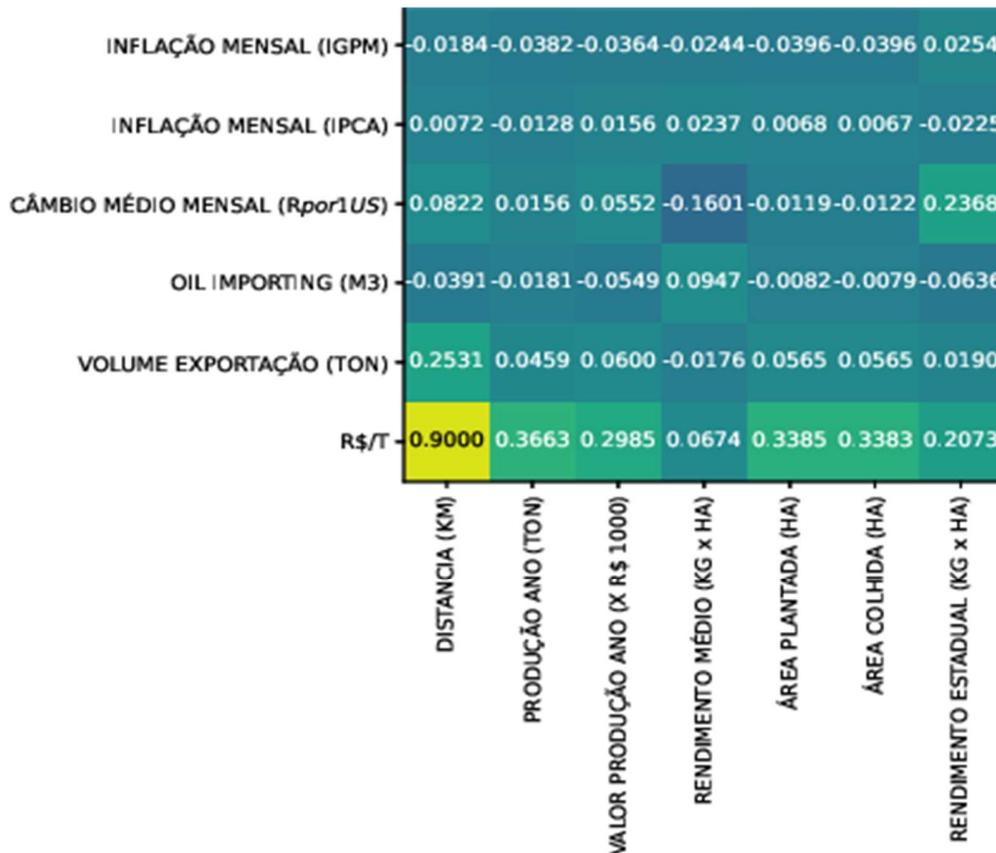


Figura 4.7 – Matriz de correlação de correlação de Pearson

Já no âmbito dos valores previstos pelo modelo, a Figura 4.8 apresenta os gráficos de dependência parcial para as variáveis que apresentaram a relação mais forte para com a variável Preço do Frete previsto, as mesmas que figuraram entre os componentes selecionados pela aplicação da técnica PCA executada anteriormente. Podemos identificar novamente a dependência da variável Distância (km) na previsão do frete. Por outro lado, as demais variáveis detêm uma relação menor quando em comparação com a Distância (km).

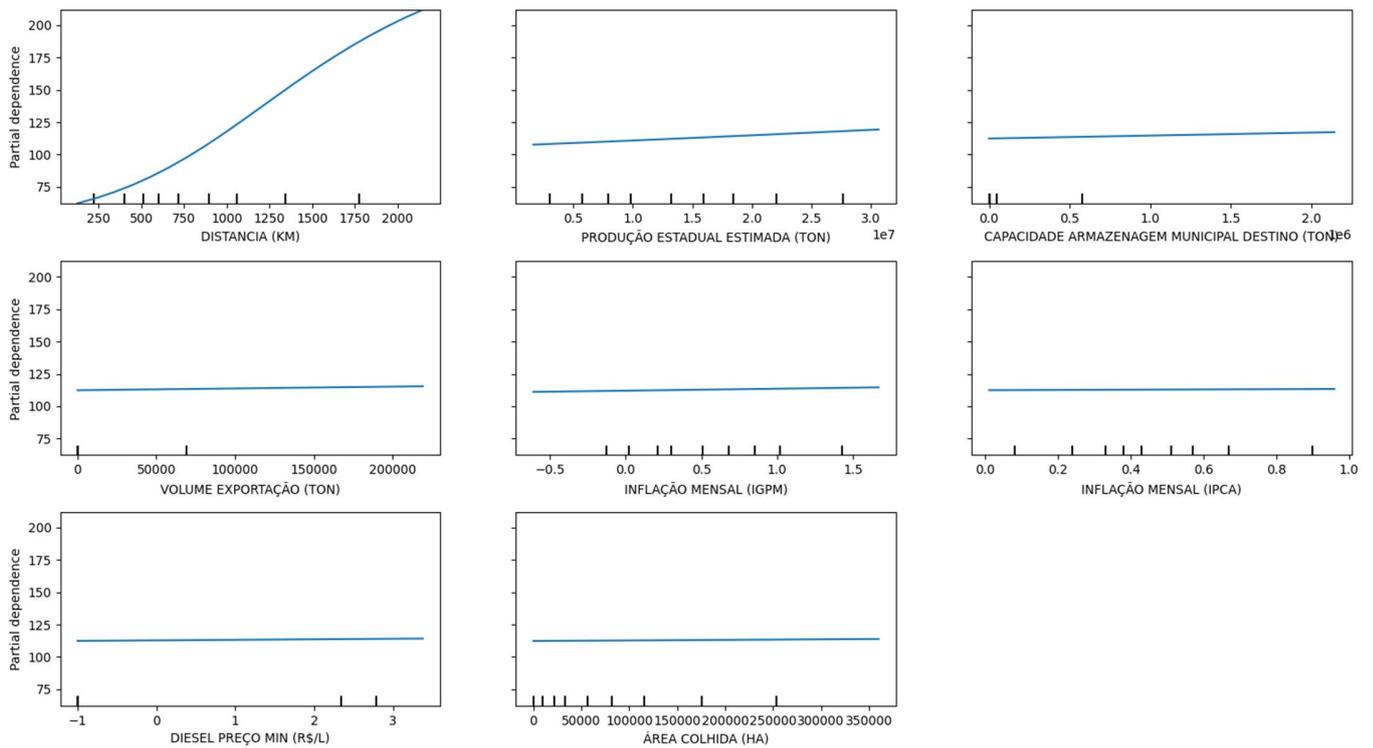


Figura 4.8 – Partial Dependence Plot para o Preço do Frete previsto pelo modelo.

O último passo do processo de previsão de preço das rotas de frete consiste na criação e aplicação do modelo único de MLP sobre os conjuntos de treinamento e teste. Para tal criamos o modelo com os valores dos hiperparâmetros números de neurônios na camada escondida presentes na Tabela 4.1, e penalizações presentes na Tabela 4.3. Criamos assim dois modelos, cada um deles configurados com um dos valores de penalização identificados pela técnica de validação cruzada: 0,1, que otimiza as métricas para rotas curtas, e 100,0, que otimiza as métricas para rotas longas (Tabela 4.4), aplicados sobre os conjuntos de treinamento e teste.

Tabela 4.4 – Resultados dos dois modelos de MLP única

Otimização	L2	Neurônios	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
MLP rotas curtas	0,1	91	0,8418	9,7448	0,9473	14,9264
MLP rotas longas	100,0	91	0,5093	17,1649	0,8279	26,9831

O modelo parametrizado com penalização igual a 0,1 (Tabela 4.4) resultou em um RMSE de 9,7448, e  $R^2$  de 0,8418 para o conjunto das rotas curtas, e RMSE 14,9264, e  $R^2$  de 0,9473, para as rotas longas. Com isso, apesar do resultado para as rotas curtas ter se revelado muito próximo do almejado ( $R^2$  alvo: 0,85), o que pode ser constatado através da análise visual da Figura 4.9, em que os

valores de frete previstos pelo modelo (pontos vermelhos) sobrepõem grande parte dos valores reais (pontos azuis), as métricas para as rotas longas indicam a presença de um sobre-ajuste.

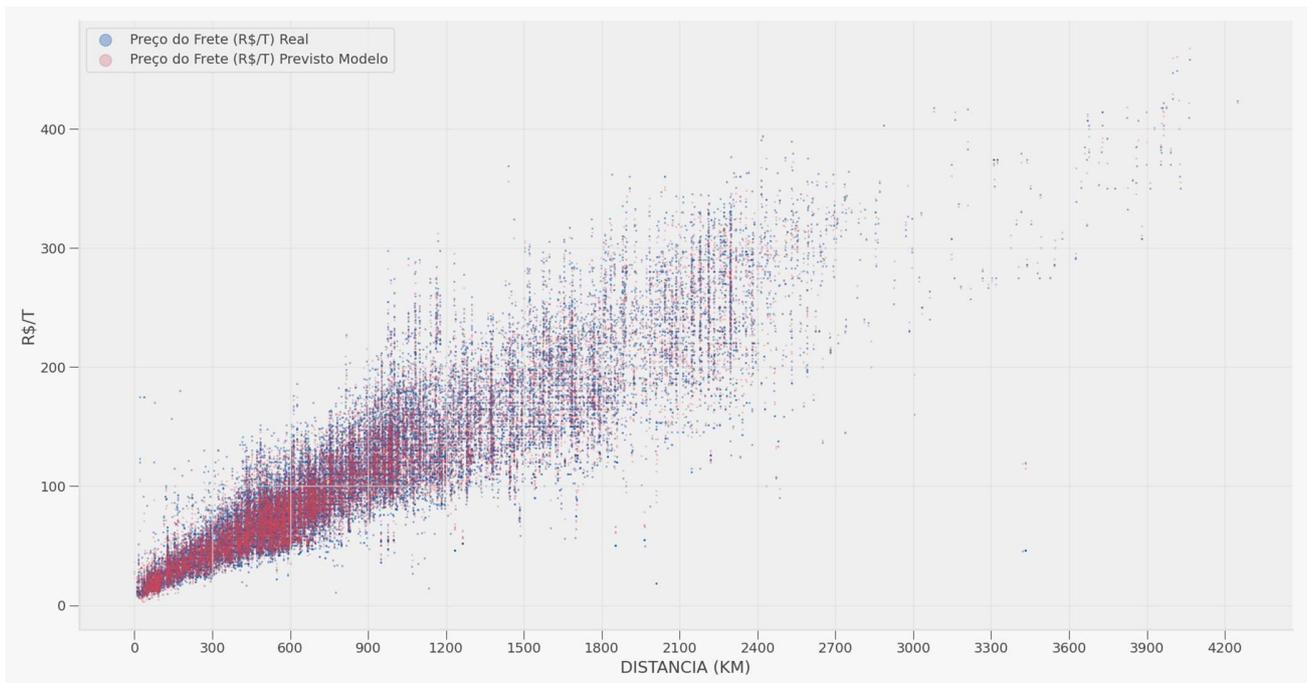


Figura 4.9 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização *Ridge Regression* igual a 0,1.

O último modelo analisado nessa seção, que trata de um único modelo de MLP aplicado à previsão de preços de fretes em ambas as partições das rotas, foi parametrizado com a penalização igual à 100,0, com o objetivo de reduzir o sobre-ajuste para rotas mais longas. Com isso obtivemos um valor de 17,1649 para o erro, e 0,5093 para o  $R^2$ , das rotas curtas, e 26,9831, e 0,8279 das rotas longas. Dessa vez, apesar do sobre-ajuste ter sido mitigado para as rotas longas, passamos a ter o problema simetricamente oposto nas rotas curtas, uma vez que o baixa  $R^2$  revela uma tendência de *underfitting*. Esse ponto pode ser intuitivamente percebido através da análise visual da Figura 4.10, através do qual apresentamos o comportamento dos valores previstos por esse modelo junto aos valores reais. Podemos identificar que o intervalo das rotas até 600km sofreu em demasia com a penalização, à medida que uma área de aproximadamente metade dos valores reais não coincide com os valores previstos, o que vai ao encontro das métricas obtidas pelo modelo.

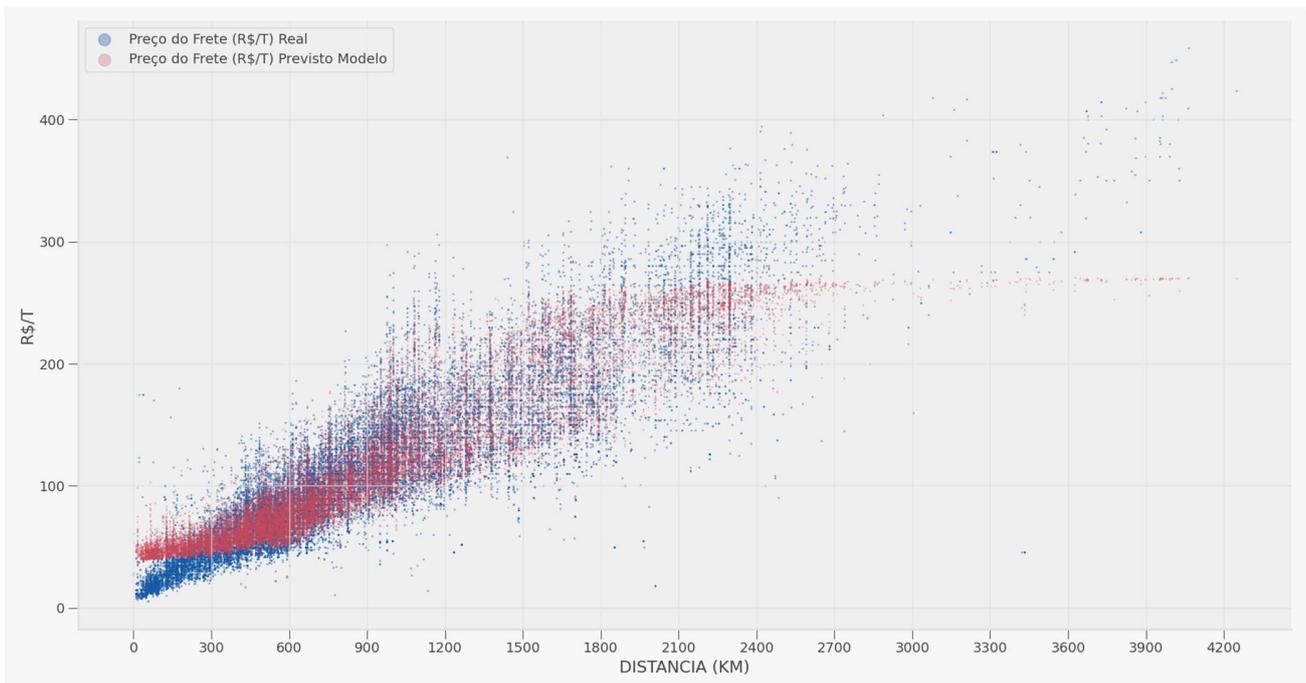


Figura 4.10 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização *Ridge Regression* igual a 100,0.

#### 4.1.2 MLPs Individuais – Dois Modelos para Duas Partições

Nessa etapa do trabalho adotamos dois modelos de MLP parametrizados e treinados de maneira independente, em contraposição ao processo do modelo único descrito na seção anterior. Cada um dos dois modelos é exclusivamente designado para tratar de uma das partições dos dados – rotas iguais ou mais curtas que 600km, ou rotas mais longas que 600km – não havendo dessa forma nenhuma interferência nos dados utilizados pelo outro modelo, ou nos resultados gerados por cada um dos modelos.

A técnica de validação cruzada *Hold-out* foi novamente utilizada com o objetivo de descobrir o número ideal de neurônios a ser empregado na camada escondida da rede. Na primeira execução objetivamos a descoberta daquele hiperparâmetro para o modelo destinado a tratar as rotas curtas. Nesse caso a técnica retornou 84 (Figura 4.11) como número ideal de neurônios, pois através desse valor obtivemos o par de métricas,  $R^2$  igual a 0,8270 e RMSE igual a 10,0756, que representaram o melhor resultado para essa partição dos dados.

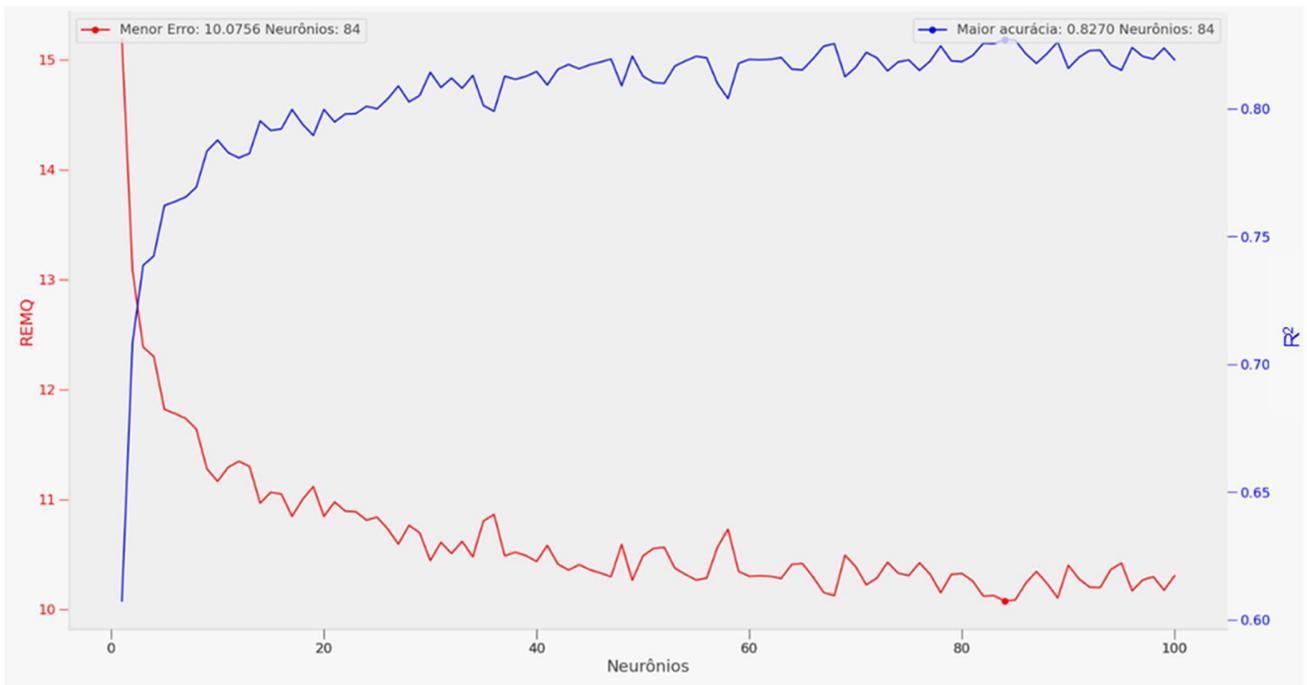


Figura 4.11 – RMSE,  $R^2$  e número ideal de neurônios do conjunto distância  $\leq 600$  km, com *Hold-out* e modelo dedicado.

Na segunda execução da técnica de validação cruzada os conjuntos de treinamento e validação se restringiram à partição das rotas longas, o que resultou em um valor de 100 para o número de neurônios da cama escondida (Figura 4.12), sob um  $R^2$  de 0,9320, e RMSE de 16,9457.

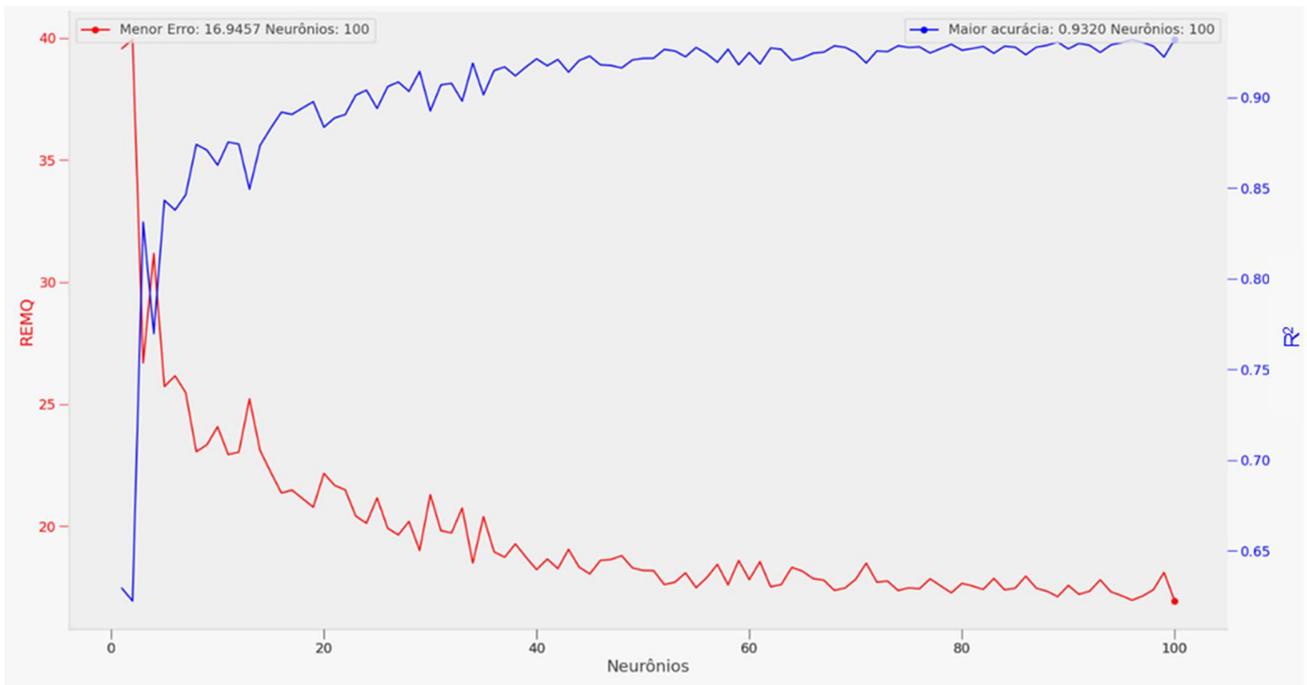


Figura 4.12 – RMSE,  $R^2$  e número ideal de neurônios do conjunto distância > 600 km, com *Hold-out* e modelo dedicado.

Semelhantemente aos resultados que apresentamos na seção anterior, obtidos através da utilização de um modelo único de MLP, temos novamente dois resultados diferentes para o hiperparâmetro em questão, obtidos a partir das duas execuções da técnica. Todavia, uma vez que estamos fazendo uso de dois modelos diferentes ao invés de um único, podemos selecionar aqueles que apresentaram melhor resultado em seus respectivos domínios dos dados, sem o ônus dos resultados potencialmente indesejados na outra partição dos dados. Em resumo, o uso de dois modelos ao invés de um só nos permite, nesse ponto, usar as duas configurações de MLP, uma com 84 neurônios, e outra com 100 neurônios, de maneira simultânea, e não exclusivamente uma das duas, como no caso do modelo único.

A seguir, da mesma forma que procedemos para com o modelo único de MLP, empregamos a técnica de PCA com objetivo de analisar uma possível redução da dimensionalidade sobre os dados. Através do *Hold-out* buscando a variância que resultasse nos melhores valores para as métricas, identificamos uma variância de 0,93 (Tabela 4.5), que levou a um  $R^2$  de 0,8262, RMSE de 10,0991 e 15 componentes selecionados, no modelo das rotas curtas; e uma variância de 0,94, que levou a um  $R^2$  de 0,9309, RMSE de 17,0599 e 16 componentes, no modelos aplicado às rotas longas. Quando comparadas as métricas resultantes do PCA percebemos que ficaram levemente inferiores do que aquelas obtidas na busca pelo hiperparâmetro quantidade de neurônios da camada escondida (Figura 4.11 e

Figura 4.12). Assim, uma vez mais, optamos por descartar o uso do PCA, e seguir com todas as colunas do conjunto completo de dados.

Tabela 4.5 – Resultados do *Hold-out* para busca pela variância ideal

Otimização	Neurônios	Variância	$R^2$	RMSE
MLP rotas curtas	84	0,93	0,8262	10,0991
MLP rotas longas	100	0,94	0,9309	17,0599

Como executado anteriormente no contexto do modelo único de MLP, aplicamos mais uma vez a técnica de validação cruzada *Hold-out* com objetivo de descobrir o valor mais adequado ao hiperparâmetro L2, que representa a penalização ao modelo. Esse processo permite que o sobre-ajuste do modelo seja reduzido, o que, em nosso caso, tem ocorrido majoritariamente no modelo que trata das rotas mais longas, conforme pudemos observar na Figura 4.12, em que o  $R^2$  chegou a 0,9320.

O modelo aplicado às rotas curtas, e parametrizado com 84 neurônios na camada escondida, retornou uma penalização de 0,1 (Figura 4.13), levando a um erro de 10,0635 e  $R^2$  0,8275. Dessa forma, apesar de quase que anular a penalização, ainda assim o modelo não foi capaz de atingir o  $R^2$  alvo pré-definida (0,85).

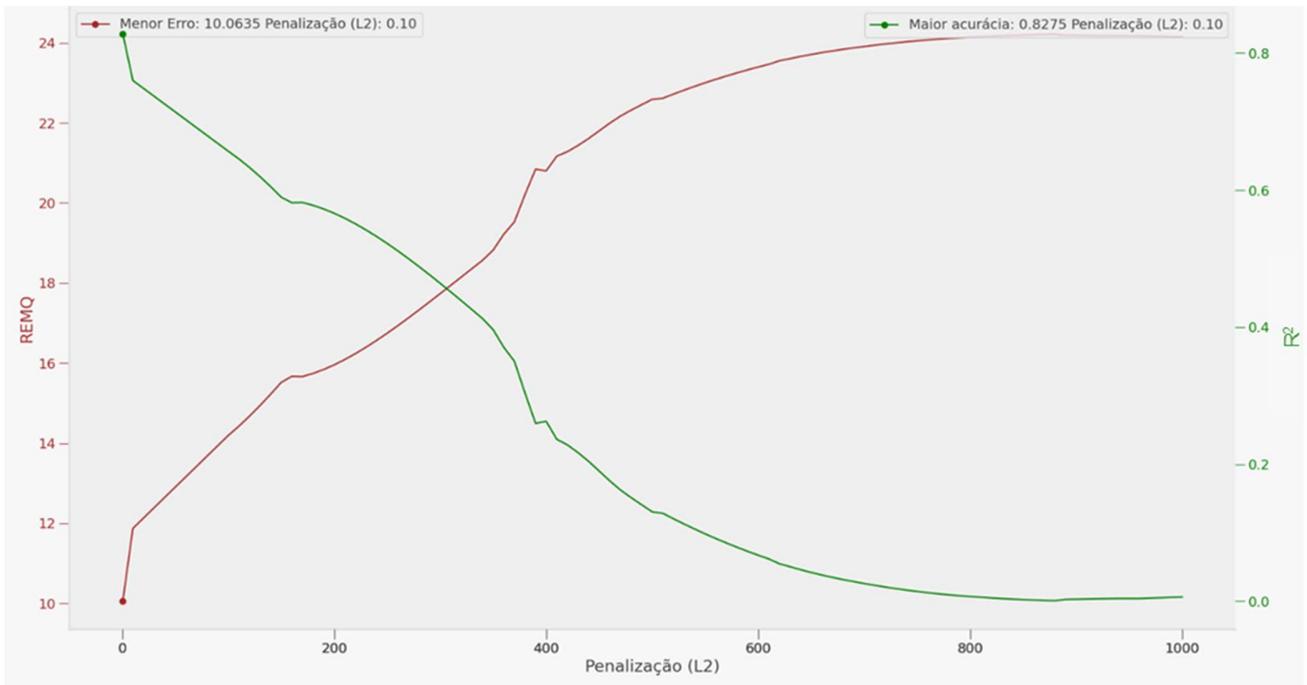


Figura 4.13 –  $RMSE$  e  $R^2$  do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.

Analogamente, para descobriremos o valor para o hiperparâmetro da penalização que leve ao melhor resultado, executamos o mesmo procedimento para o subconjunto das rotas longas. O número de neurônios da camada escondida foi configurado para 100, conforme retornado pela *Hold-out* empregado na pesquisa daquele hiperparâmetro (Figura 4.12). Dessa forma o valor retornado para a penalização foi 100,0, que levou a um  $R^2$  de 0,8459 e erro de 25,4659 (Figura 4.14). Esse resultado revelou-se mais próximo do  $R^2$  alvo em comparação aquele gerado pelo modelo único (0,8390, como vimos na Figura 4.6), quando aplicado à mesma partição dos dados.

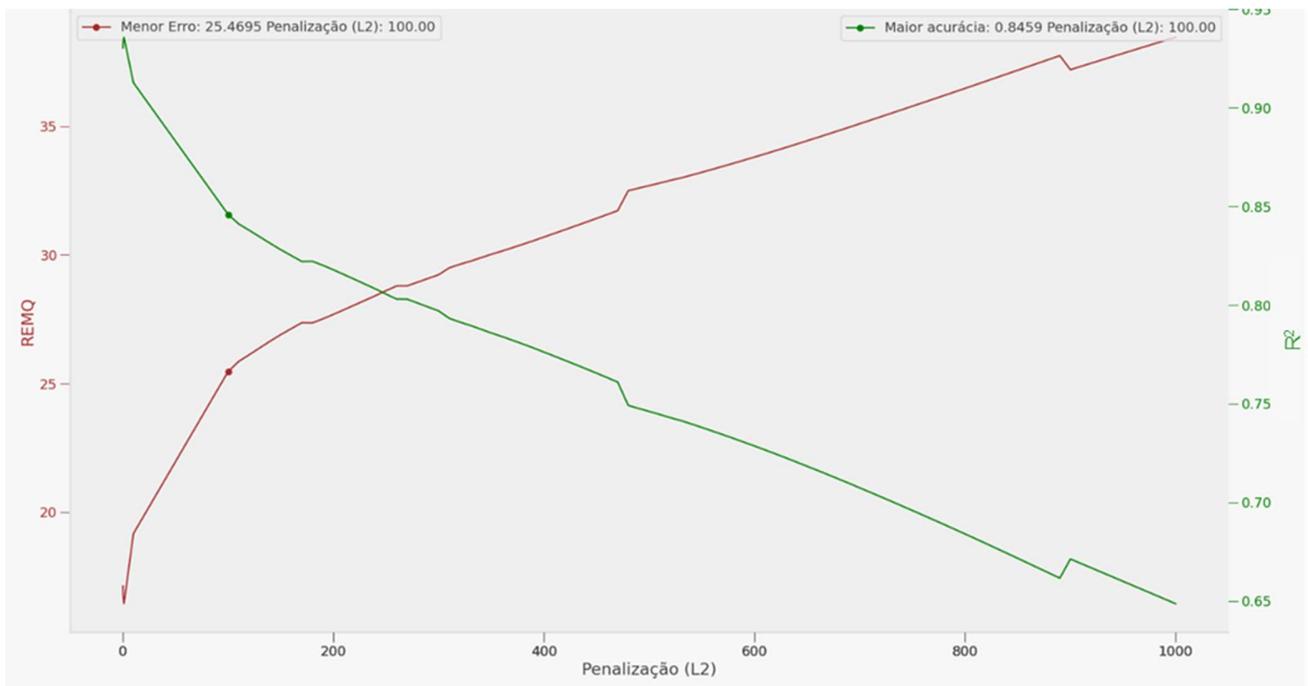


Figura 4.14 –  $RMSE$  e  $R^2$  do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.

Após a obtenção dos valores ideais para os hiperparâmetros, número de neurônios na camada escondida e penalização L2, para ambos os modelos que respectivamente foram criados para prever o preço dos fretes para rotas curtas (iguais ou menores que 600km), e longas (maiores que 600km), passamos à etapa da predição dos fretes a partir dos conjuntos de treinamento e teste.

De fato, a utilização de um modelo de MLP, configurado e executado de maneira independente, para cada partição dos dados, revelou-se mais flexível que a adoção do modelo único, à medida que cada um dos dois modelos pode ser treinado e parametrizado individualmente. Na Tabela 4.6 percebemos que a possibilidade de refinar os hiperparâmetros de cada modelo, e posteriormente treiná-los sobre cada partição individual, permitiu uma especialização sobre cada subconjunto dos dados, sem a obrigatoriedade da previsão de valores na outra partição. Assim, nessa etapa, pudemos comparar as métricas da abordagem de modelos individuais, com aquelas do modelo de MLP única, resultando em um  $R^2$  de 0,8785 para as rotas curtas, da primeira abordagem, em comparação ao valor de 0,8418 (Tabela 4.4) retornado pela segunda; além de 0,8506 para as rotas longas, também como resultado das MLPs individuais, em comparação a 0,8279 da MLP única.

Tabela 4.6 – Resultados dos dois modelos de MLP independentes, aplicados sobre os conjuntos de treinamento e teste.

Otimização	L2	Neurônios	$R^2$	RMSE
MLP rotas curtas	0,1	84	0,8785	8,5404
MLP rotas longas	100,0	100	0,8506	24,8818

Os resultados retornados pela abordagem das MLPs individuais foram inegavelmente melhores, pois ficaram mais próximos do  $R^2$  alvo, sob um erro mais baixo. Todavia a análise do gráfico de dispersão da Figura 4.15 revelou uma faixa de pontos azuis no intervalo das rotas de 600km a 900km. Aquele conjunto de pontos azuis (reais) sem sobreposição de pontos vermelhos (previstos) intuitivamente significa presença de erro. De fato, ao analisarmos as métricas naquele intervalo isoladamente encontramos um RMSE de 18,8371 e  $R^2$  de 0,4869. Esses valores tão distantes do esperado ocorreram em função da penalização empregada sobre o modelo que trata as rotas mais longas, pois o erro elevado e queda no  $R^2$  tem início abrupto justamente no começo do intervalo de rotas tratado pelo segundo modelo.

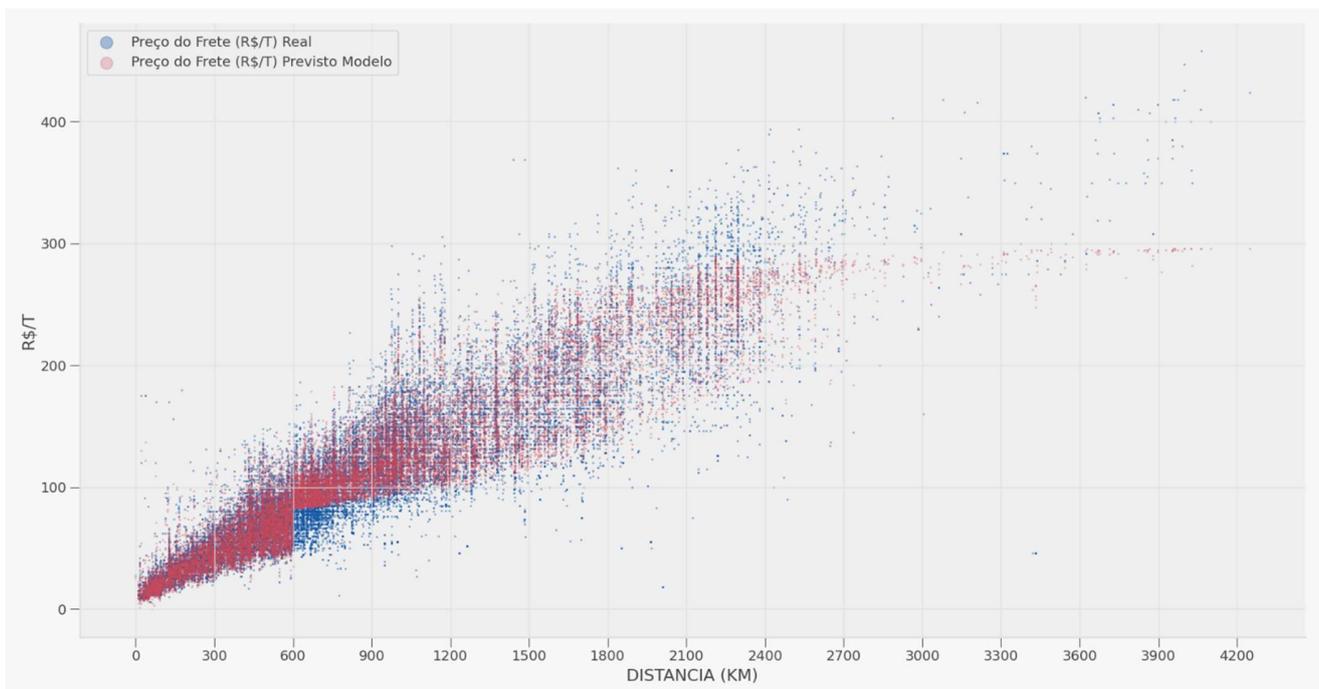


Figura 4.15 – Dispersão dos valores previstos pelos modelos individuais para rotas curtas e longas, e dos valores reais, com regularização *Ridge Regression* iguais a 0,1 e 100,0 respectivamente.

## 4.2 Resultados da Aplicação da SVR

A utilização da técnica de SVR para previsão de preço de rotas de frete esteve prevista no escopo inicial desse trabalho, todavia ela se torna especialmente importante como alternativa viável, uma vez que todas as abordagens exploradas no uso da técnica MLP apresentaram pontos negativos, conforme vimos na seção anterior.

O roteiro seguido para aplicação da técnica SVR foi o mesmo utilizado para a MLP. Sendo assim essa fase foi dividida em duas etapas, a primeira em que um único modelo de SVR é empregado para prever o preço das rotas de frete, hora focado na otimização das rotas mais curtas, hora considerando apenas as métricas oriundas das rotas mais longas. Essa abordagem é descrita na seção 4.2.1. Na segunda etapa a restrição do modelo único é removida, e passamos a contar com um modelo de SVR específico operando exclusivamente, e de maneira simultânea, sobre cada partição. Essa segunda abordagem é explicada na seção 4.2.2. Em ambas as etapas, tal qual no caso da MLP, os passos foram rigorosamente os mesmos: primeiro adotamos processos de validação cruzada (*Hold-out*) para descoberta dos valores ideais dos hiperparâmetros do modelo, depois a mesma técnica é adotada na busca do valor da penalização, e finalmente avaliamos os resultados sobre os conjuntos de treinamento e teste.

### 4.2.1 SVR Única – Um Modelo para Duas Partições

No uso kernel gaussiano, adotado por esse trabalho, o hiperparâmetro *gamma* consiste no grau de influência que uma amostra do treinamento possui. Quanto maior o valor de *gamma*, mais próximas outros pontos necessitam estar para que sejam afetados por uma amostra individual (GUENTHER; SCHONLAU, 2016). Dessa forma é crucial que o valor assumido por *gamma* seja aproximado através de uma técnica de validação cruzada.

Para que o valor ideal de *gamma* pudesse ser encontrado através da técnica *Hold-out*, o hiperparâmetro variou ao longo das execuções em um intervalo de 0,01 a 1,00, em incrementos de 0,01. Assim, quando o foco foi a otimização das métricas da partição das rotas curtas, o valor de *gamma* que retornou as melhores métricas foi 0,01 (Figura 4.16), que levou a um  $R^2$  de 0,4722 e RMSE e 17,6567. Esse resultado, apesar de ser o melhor dentre todas os modelos disparados pela validação cruzada, ainda

é baixo quando em comparação com as métricas retornadas pela etapa análoga da técnica de modelo único de MLP, que se deu na busca do valor ideal para os neurônios da camada escondida (Tabela 4.1). Isso ocorre porque ainda nessa etapa estamos trabalhando com uma penalização forte para o SVR. Isso é necessário para fins de performance e melhor uso do tempo, pois penalizações menos fortes na biblioteca utilizada pela implementação desse trabalho, resultam em um longo tempo de execução, o que poderia dificultar a geração dos resultados. De qualquer maneira a penalização será exercitada nas etapas posteriores desse fluxo de trabalho, quando então as métricas assumem valores semelhantes aos apresentados na seção em que apresentamos os resultados da MLP.

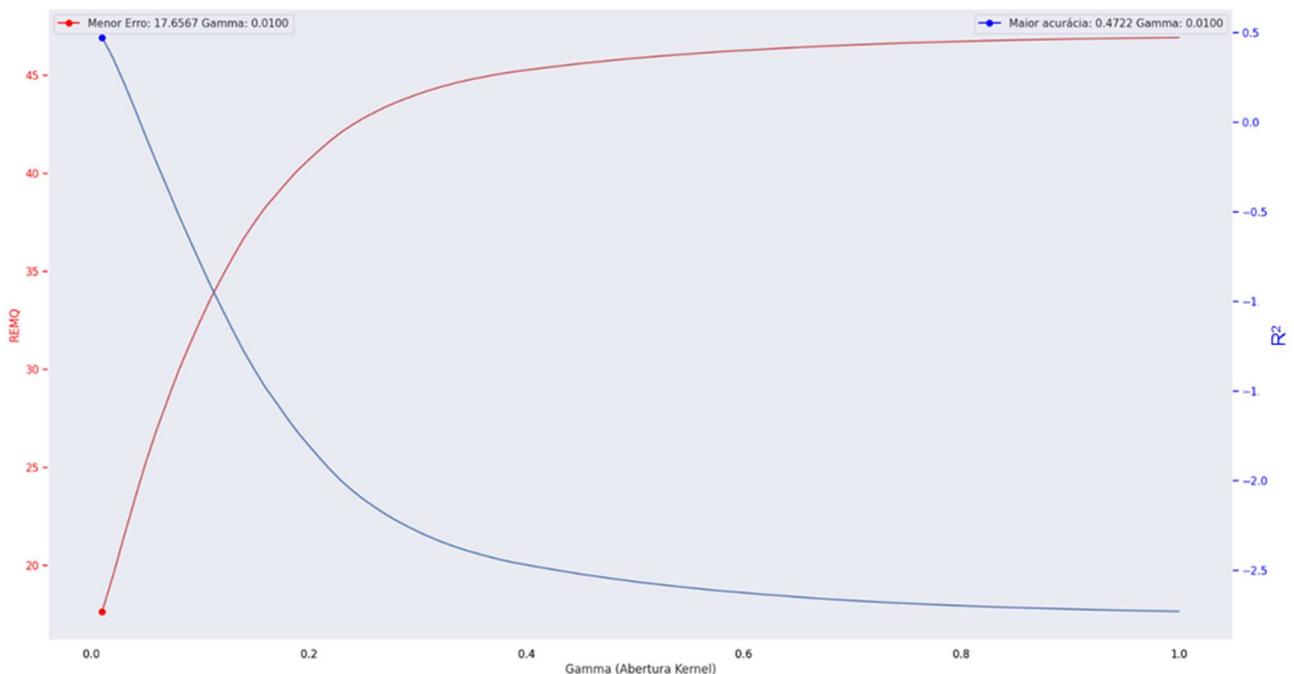


Figura 4.16 – RMSE e  $R^2$  do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição.

A segunda execução do *Hold-out*, que teve como foco as métricas da partição das rotas longas, na busca do valor ideal para  $\gamma$ , chegou ao mesmo resultado para o hiperparâmetro: 0,01 (Figura 4.17), com  $R^2$  de 0,7650, e RMSE de 30,9358.

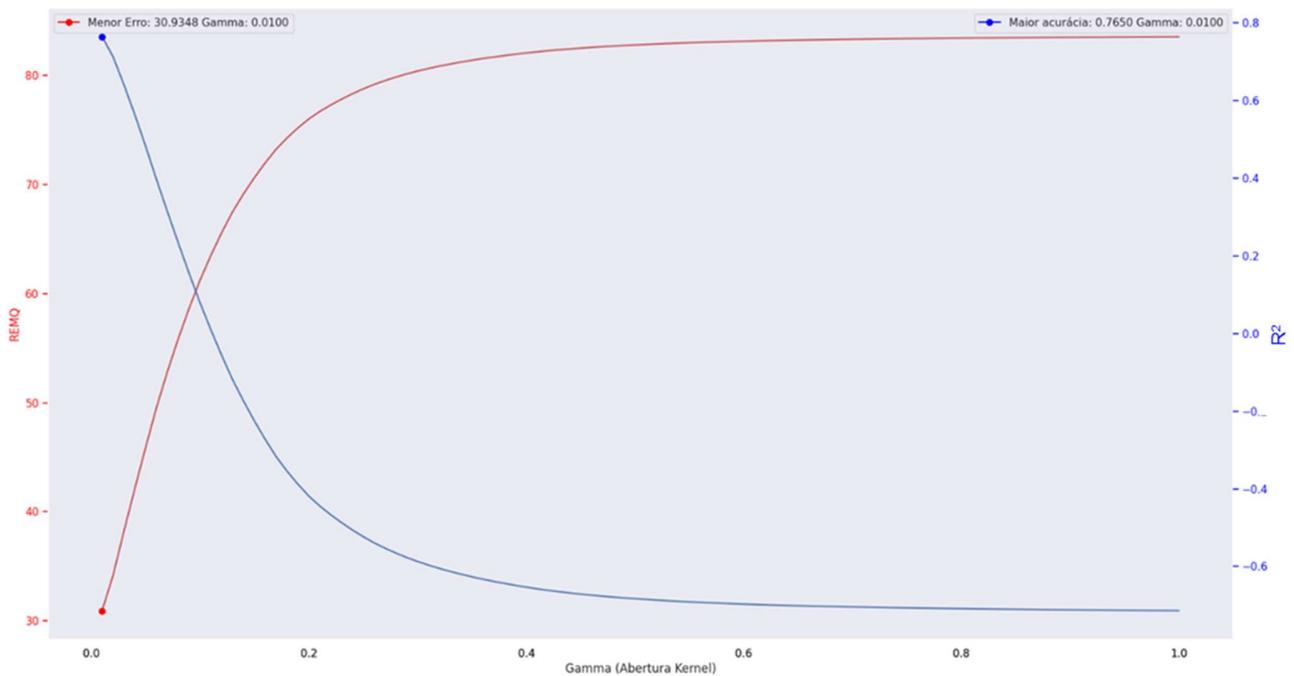


Figura 4.17 – RMSE e  $R^2$  do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição.

Quando tabulamos e comparamos os resultados de ambas as execuções da técnica de validação cruzada, conforme podemos visualizar na Tabela 4.7, percebemos que diferentemente da técnica de MLP única, o *Hold-out* quando utilizado com o SVR, para nossos conjuntos de dados, retornou um único e comum valor para hiperparâmetro em questão (0,01). Além disso, as métricas obtidas foram precisamente as mesmas, independente da partição dos dados avaliado pela técnica. Esse resultado elimina, ao menos na etapa da busca pelo *gamma*, o problema enfrentado pela MLP única, em que obtivemos diferentes quantidades de neurônios quando o foco de otimização ocorreu em diferentes partições dos dados, e levou a sobre-ajuste e sub ajuste em uma das partições, dependendo da configuração analisada.

Tabela 4.7 – Resultados do *Hold-out* aplicado na busca pelo valor de *Gamma* ideal

Otimização	<i>Gamma</i>	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
SVR rotas curtas	0,01	0,4722	17,6567	0,7650	30,9348
SVR rotas longas	0,01	0,4722	17,6567	0,7650	30,9348

Da mesma forma que procedemos com o MLP, empregamos a técnica PCA em conjunto com a validação cruzada e o SVR, para avaliar uma possível redução da dimensionalidade nos dados, e seu

efeito sobre o comportamento das métricas. Como antes, fizemos a variância assumir valores entre 0,01 e 0,99 para que o *Hold-out* retornasse o número ideal para o melhor par de métricas. Assim obtivemos uma variância de 0,98 (Tabela 4.8), que levou a um  $R^2$  de 0,4783 e erro de 17,5541 para as rotas curtas; e  $R^2$  de 0,7648 e erro de 30,9433 para as rotas longas, com o mesmo valor de variância. Essa configuração levou a seleção de 16 componentes, sendo eles: rendimento médio (kg/ha), valor da produção no ano (multiplicado por R\$ 1.000,00), capacidade armazenagem municipal origem (t), distância (km), gasolina preço mínimo (R\$/l), mercado internacional Chicago (US\$/dia), mercado internacional paridade (R\$/saca 60kg), armazéns no município de destino, capacidade instalada indústria moagem no município de destino (t/dia), mês, armazéns no município de origem, capacidade armazenagem municipal destino (t), capacidade instalada indústria moagem no município de origem (t/dia), commodity, destino, atacado (R\$/saca 60kg). Apesar dos valores de  $R^2$  e erro ficarem muito próximos dos resultados das métricas na etapa da busca do *gamma* (Tabela 4.7), optamos por manter a totalidade das colunas, por compreendermos que as performances das execuções não variaram drasticamente mesmo com a configuração sugerida pelo PCA.

Tabela 4.8 – Resultados do *Hold-out* aplicado na busca pela variância ideal, para uso em PCA

Otimização	Variância	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
SVR rotas curtas	0,98	0,4783	17,5541	N/A	N/A
SVR rotas longas	0,98	N/A	N/A	0,7648	30,9433

A busca pelo valor da penalização, como no caso da MLP, se deu através da validação cruzada. Como mencionamos brevemente no início dessa seção, a penalização, também conhecida por hiperparâmetro  $C$ , na biblioteca utilizada por nossa implementação, é vista como o quadrado de  $L_2$ , e sua força é inversamente proporcional ao valor designado para  $C$ . Isso significa que valores mais elevados resultam em uma penalização mais fraca, *i.e.*, enquanto valores altos forçam a classificação correta dos dados, valores mais baixos resultam em uma margem mais suave. Ainda, valores mais baixos resultam em modelos mais performáticos (menor tempo de treinamento), e por isso optamos por eles até esse ponto, no procedimento aplicação da técnica SVR.

Com objetivo de evitar sobre-ajuste, definimos o  $R^2$  alvo como 0,85 para a busca do hiperparâmetro  $C$ , que foi o mesmo valor adotado para a penalização da MLP. Para encontrar o valor

ideal para  $C$ , *Hold-out* foi empregado em uma lista de valores para  $C$ , que variou de 0,001 a 1000,0 em passos de 10,0, que até 100,0 foram multiplicações, e a partir de 100,0 foram adições. Assim, o valor ideal identificado para as rotas curtas foi 120,0 (Figura 4.18), que levou a um  $R^2$  de 0,8504, e RMSE de 9,3993.

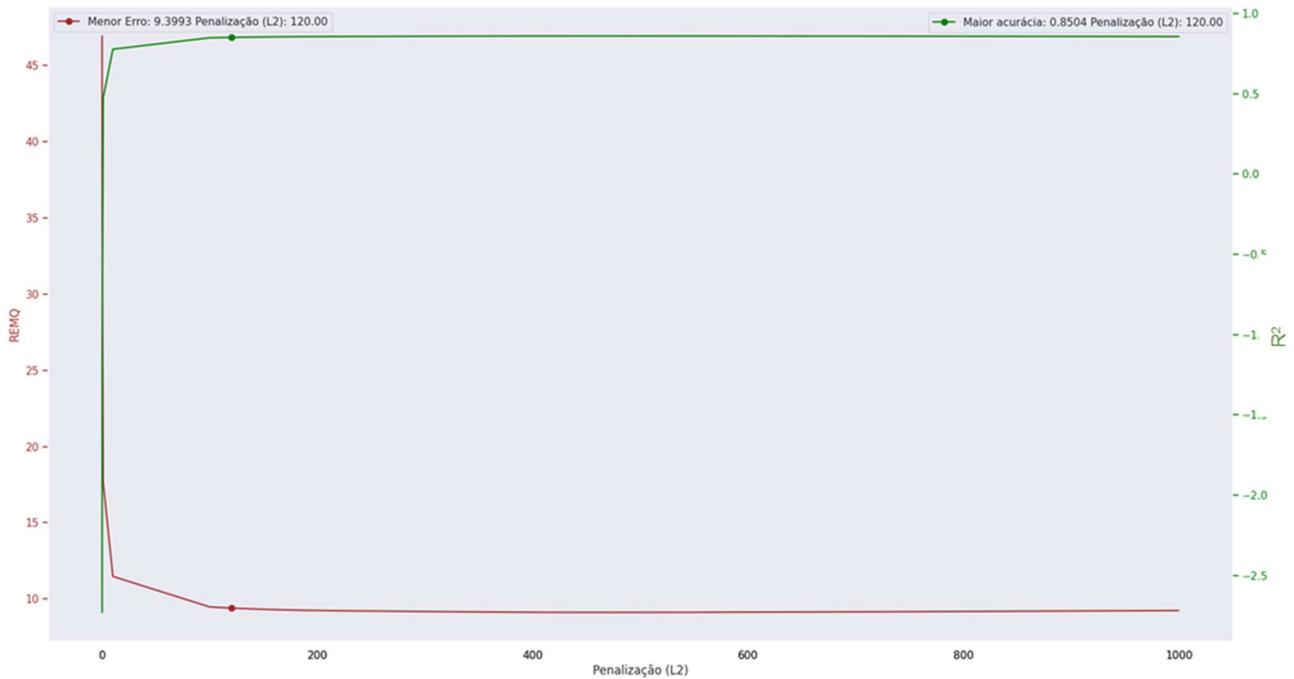


Figura 4.18 –  $RMSE$  e  $R^2$  do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.

Para as rotas mais longas, o *Hold-out* retornou o valor 10,0 (Figura 4.19) para o hiperparâmetro  $C$ , que levou a um  $R^2$  de 0,8839, e RMSE de 21,7458.

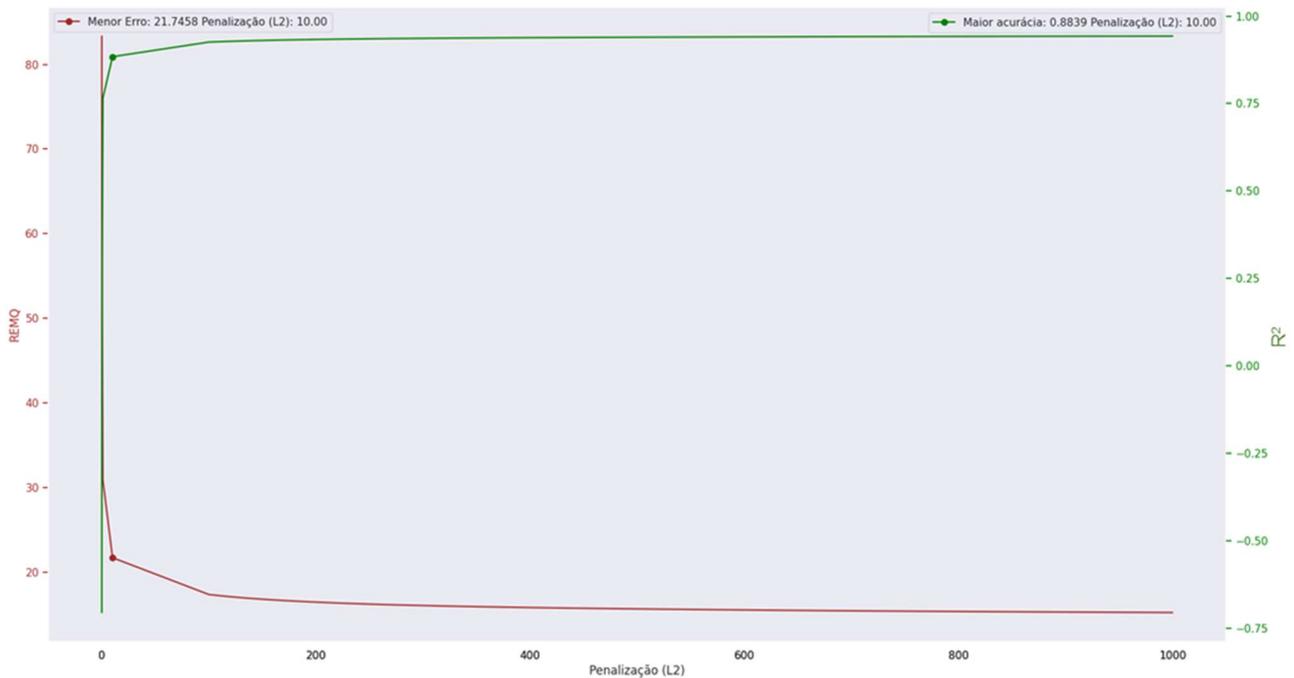


Figura 4.19 –  $RMSE$  e  $R^2$  do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização L2.

Na Tabela 4.9 podemos visualizar o resultado de ambos os modelos e suas respectivas partições de atuação. Apresentamos também na Tabela 4.9 as métricas da partição que não participou da otimização para aquele modelo, pois o modelo deve prever o preço do frete na totalidade dos dados, mesmo que seus hiperparâmetros tenham sido refinados para um subconjunto das rotas. Aqui percebemos um problema semelhante ao encontrado na investigação da MLP única, pois ao considerarmos as métricas de somente uma das partições na escolha dos valores para os hiperparâmetros, as métricas da outra partição apresentam um resultado aquém do esperado. Dessa forma, mesmo que tenhamos encontrado um resultado quase que exato para as rotas curtas, uma vez que o  $R^2$  0,8504 coincide com o  $R^2$  alvo, o efeito colateral foi um  $R^2$  elevada (0,9279) quando o mesmo modelo tenta prever as rotas longas, revelando uma tendência a sobre-ajuste. Por outro lado, quando o foco da otimização são as rotas longas, mesmo o  $R^2$  (0,8839) estando bem próxima do (valor) alvo, a mesma métrica em relação às rotas curtas encontra-se abaixo do esperado (0,7770), novamente levando a um cenário de sub-ajuste.

Tabela 4.9 – Resultados do *Hold-out* aplicado na busca pelo valor da penalização (L2) ideal.

Otimização	L2	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
SVR rotas curtas	120,0	0,8504	9,3993	0,9279	17,1373
SVR rotas longas	10,0	0,7770	11,4763	0,8839	21,7458

Uma vez que dispomos dos valores de todos os hiperparâmetros (*gamma* e penalização), e a análise da redução de dimensionalidade foi realizada, podemos criar e submeter o modelo aos conjuntos de treinamento e teste para então analisarmos o resultado final da abordagem de uma SVR única. Como vimos anteriormente a busca pelo *gamma* resultou em um único valor, 0,01, para ambas as partições dos dados. Todavia a validação cruzada retornou um valor individual para cada partição, por isso a análise, tal qual para a abordagem da MLP única, leva em consideração dois modelos, um para cada valor de penalização.

O primeiro modelo exercitado nessa etapa é parametrizado com os valores ideais para as rotas curtas. Assim os hiperparâmetros associados são 0,01 para *gamma*, e 120,0 para a penalização. Podemos visualizar na Tabela 4.10 o resultado das métricas para esse modelo, que são  $R^2$  0,8470, erro 9,5854 para as rotas curtas; e 0,9353 de  $R^2$ , e 16,5384 para o erro, na partição das rotas longas. As métricas para as rotas curtas estão muito próximas ao esperado, todavia quando olhamos para os resultados nas rotas longas, percebemos um comportamento que indica presença de sobre-ajuste.

Tabela 4.10 – Resultados dos dois modelos de SVR única, aplicados sobre os conjuntos de treinamento e teste.

Otimização	L2	<i>Gamma</i>	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
SVR rotas curtas	120,0	0,01	0,8470	9,5854	0,9353	16,5384
SVR rotas longas	10,0	0,01	0,7781	11,5413	0,8937	21,2037

O segundo modelo é configurado com *gamma* igual a 0,01, penalização igual à 10,0, que é o valor obtido pela validação cruzada ao otimizar as métricas para as rotas longas. A partir desse modelo obtemos um  $R^2$  de 0,7781 e erro de 11,5413 para as rotas curtas, e 0,8937, 21,2037, respectivamente para as rotas longas. Mais uma vez observamos que, ao tentarmos melhorar a métrica de uma das

partições sofremos uma piora nas métricas da outra, nesse caso o incremento no  $R^2$  das rotas longas implicou em uma diminuição na mesma métrica das rotas curtas, ficando abaixo do  $R^2$  alvo.

As diferenças entre os gráficos (Figura 4.20 e Figura 4.21) resultantes dos modelos de SVR nessa seção são mais sutis quando em comparação a análise dos gráficos da MLP de modelo único. Essa dificuldade em perceber diferenças visuais encontra respaldo nas métricas obtidas apresentadas na Tabela 4.10, que estão mais próximas umas das outras, quando em comparação as métricas resultantes da MLP de modelo único (Tabela 4.4). Ainda assim podemos perceber que uma penalização mais alta (menor valor para  $C$ ), presente na Figura 4.21, leva o modelo a errar os pontos dos extremos (preços de frete mais altos e mais baixos) nos intervalos de distância. Por exemplo, na Figura 4.21 o modelo abriu mão de acertar os valores para as rotas mais caras, no intervalo entre 3.900km e 4.200km, enquanto que na Figura 4.20, em que o modelo foi configurado com menor penalização, este tentou acertar o valor para aquelas rotas. Esse é um indício de sobre-ajuste, comportamento este que queremos evitar.

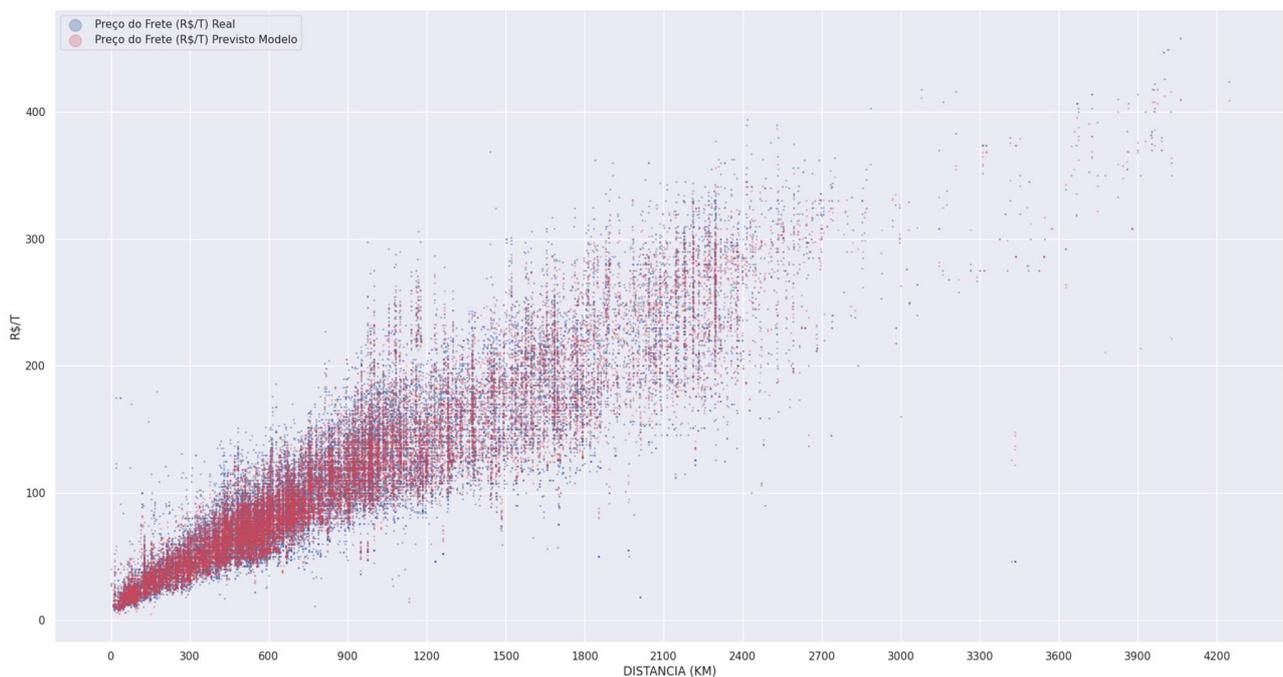


Figura 4.20 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização *Ridge Regression* igual a 120,0.

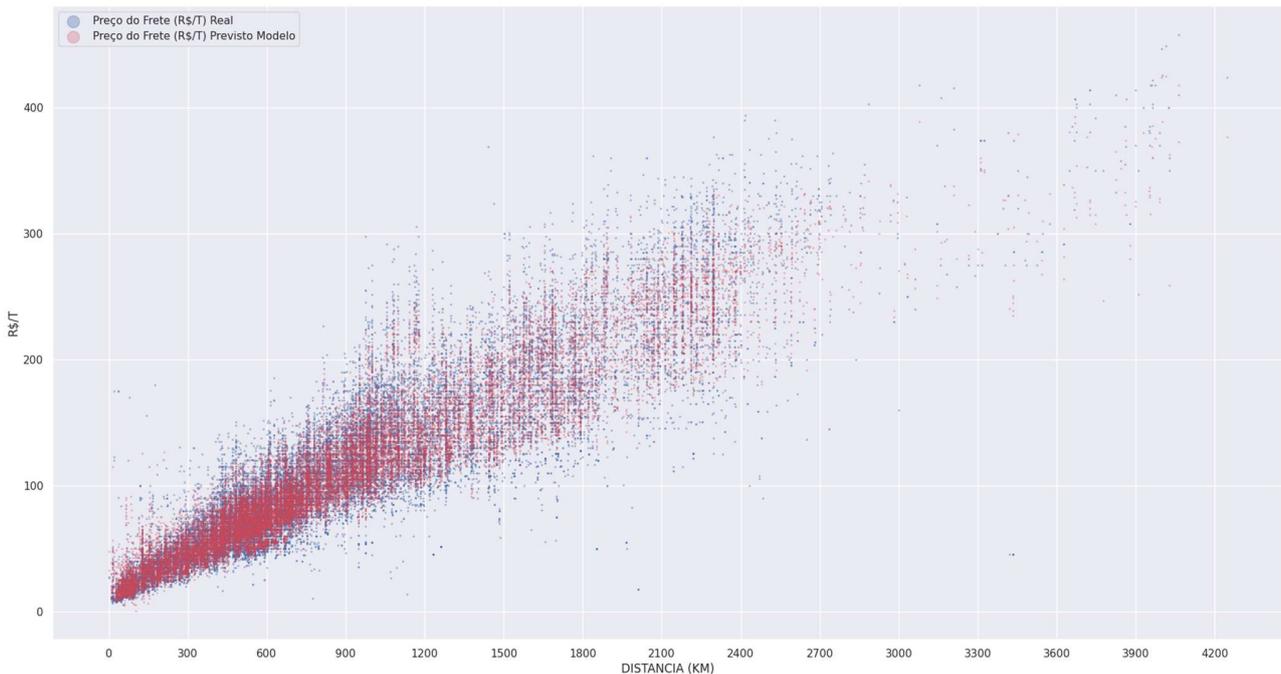


Figura 4.21 – Dispersão dos valores previstos pelo modelo e dos valores reais com regularização *Ridge Regression* igual a 10,0.

#### 4.2.2 SVRs Individuais – Dois Modelos para Duas Partições

Da mesma forma que no caso da aplicação da MLP, exercitamos também a abordagem em que um modelo de SVR, parametrizado de forma individual, trata de maneira exclusiva uma ou outra, das partições dos dados, resultando em duas SVRs especializadas.

Iniciamos novamente pela descoberta do valor ideal para o hiperparâmetro *gamma*. Aplicamos a técnica *Hold-out* sobre ambas as partições das rotas, para que, de maneira individual, o valor que resulte nas melhores métricas para a partição em análise seja retornado. Na Figura 4.22, em que projetamos a curva desempenhada pela validação cruzada para  $R^2$  e RMSE, percebemos que apesar do *gamma* ter sido variado de 0,001 até 1,0, foi logo no começo da curva, quando o hiperparâmetro assumiu o valor 0,01, que a técnica de validação cruzada retornou os melhores valores para as métricas. Nesse caso, em que a partição das rotas curtas esteve em análise, as métricas resultantes foram 0,6622 para o  $R^2$  e 14,0809 para o RMSE.

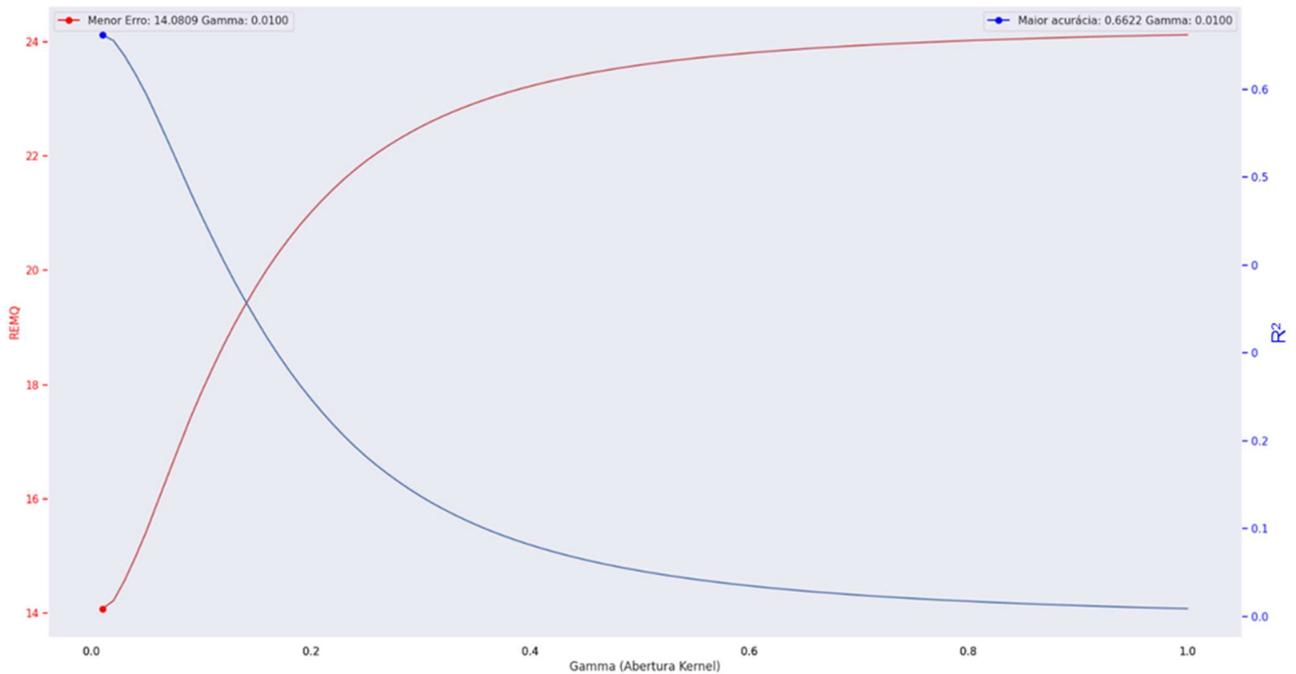


Figura 4.22 – RMSE,  $R^2$  e número ideal de neurônios do conjunto das rotas menores ou iguais a 600 km, obtidos pelo *Hold-out*, através de um modelo dedicado.

Na execução seguinte do *Hold-out* a técnica operou sobre a partição das rotas mais longas. Todavia, apesar dos dados de treinamento e validação originarem-se de um conjunto disjunto em relação à execução anterior, o valor encontrado para o hiperparâmetro  $\gamma$  foi 0,01 (Figura 4.23), o mesmo daquele encontrado para as rotas curtas. Esse valor implicou em  $R^2$  de 0,7646 e RMSE de 31,4803.

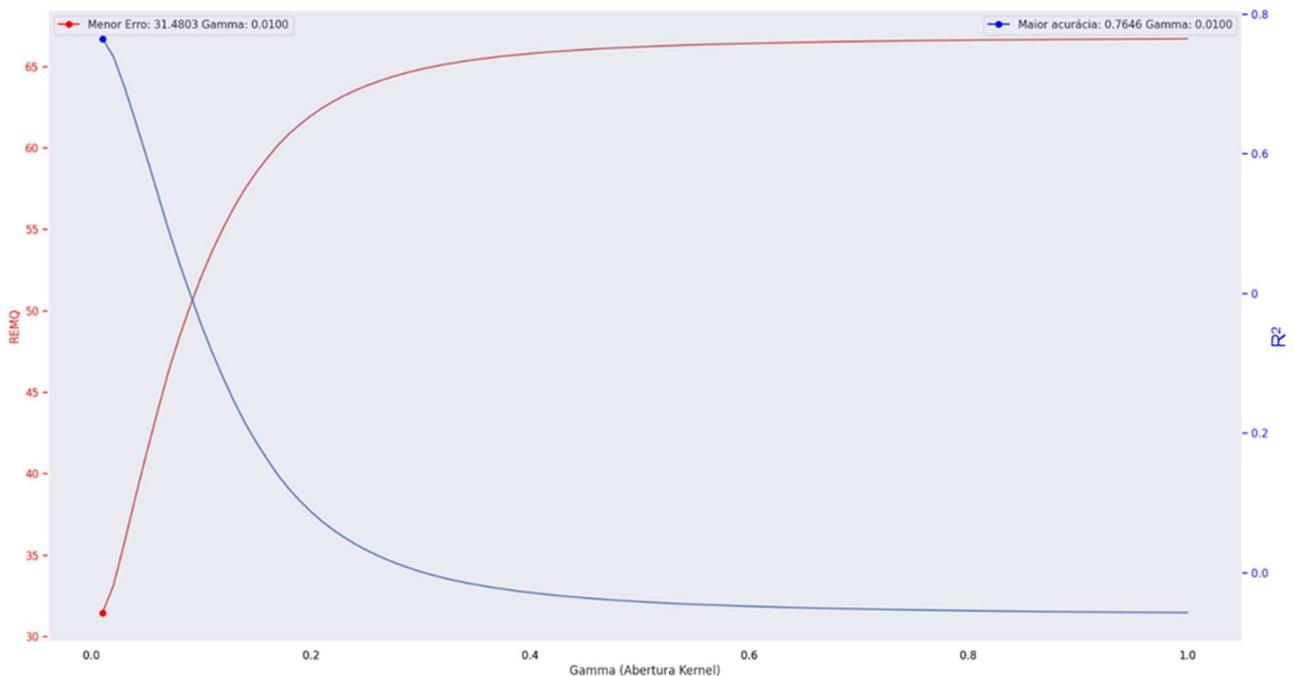


Figura 4.23 – RMSE,  $R^2$  e número ideal de neurônios do conjunto das rotas maiores que 600 km, obtidos pelo *Hold-out*, através de um modelo dedicado.

A flexibilidade do uso de modelos individuais, um especializado em cada partição, novamente se fez presente nessa etapa. Mesmo que o valor para o hiperparâmetro  $\gamma$  tenha se mantido o mesmo (0,01), a métrica  $R^2$  para as rotas curtas melhorou quando em comparação a SVR de modelo único, indo de 0,4722 para 0,6622.

A análise de redução da dimensionalidade dos dados através da técnica PCA se deu primeiramente através da validação cruzada para identificação do valor ideal para a variância. Esse passo foi executado uma vez para cada partição dos dados, sendo o hiperparâmetro  $\gamma$  configurado com o valor retornado no passo anterior (0,01 para ambas as partições). Dessa forma identificamos o valor de 0,99 (Tabela 4.11) para a variância, que para as rotas curtas resultou em um  $R^2$  de 0,6616 e RMSE de 14,0929. Para as rotas longas o *Hold-out* retornou o mesmo valor de variância, o que levou ao  $R^2$  de 0,7638 e RMSE de 31,5322. Em ambas as partições o PCA selecionou 22 componentes.

Tabela 4.11 – Resultados do *Hold-out* aplicado na busca pela variância ideal, para uso em PCA

Otimização	$\Gamma$	Variância	$R^2$	RMSE
SVR rotas curtas	0,01	0,99	0,6616	14,0929
SVR rotas longas	0,01	0,99	0,7638	31,5322

Apesar da redução considerável no número de componentes, de um total de 49 para 22 selecionados pela técnica, optamos por permanecer trabalhando com o número total de componentes, por entender que não houve mudança na performance do modelo. Outro motivo é a redução na qualidade das métricas, que apesar de pequena, fez-se presente.

Finalizada a análise dos componentes através do PCA, passamos a busca pelo valor ideal para a penalização. É nesse passo que as métricas começam a se aproximar de valores mais semelhantes aos padrões esperados, logo que, como mencionamos anteriormente, por razões de performance optamos por manter a penalização alta nas primeiras etapas, mas sem qualquer prejuízo à busca pelos valores ideais.

A aplicação da validação cruzada restrita às rotas mais curtas retornou um valor de 560,0 (Figura 4.24) para a penalização, o que levou a um  $R^2$  de 0,8333 e erro de 9,8918. Esse resultado apresenta uma leve piora em relação ao modelo único de SVR, que aplicado às rotas menores ou iguais à 600km, retornou um  $R^2$  de 0,8504. Todavia devemos recordar que esse ganho do modelo único traz a desvantagem do sobre-ajuste ( $R^2$  de 0,9279) nas rotas mais longas que 600km.

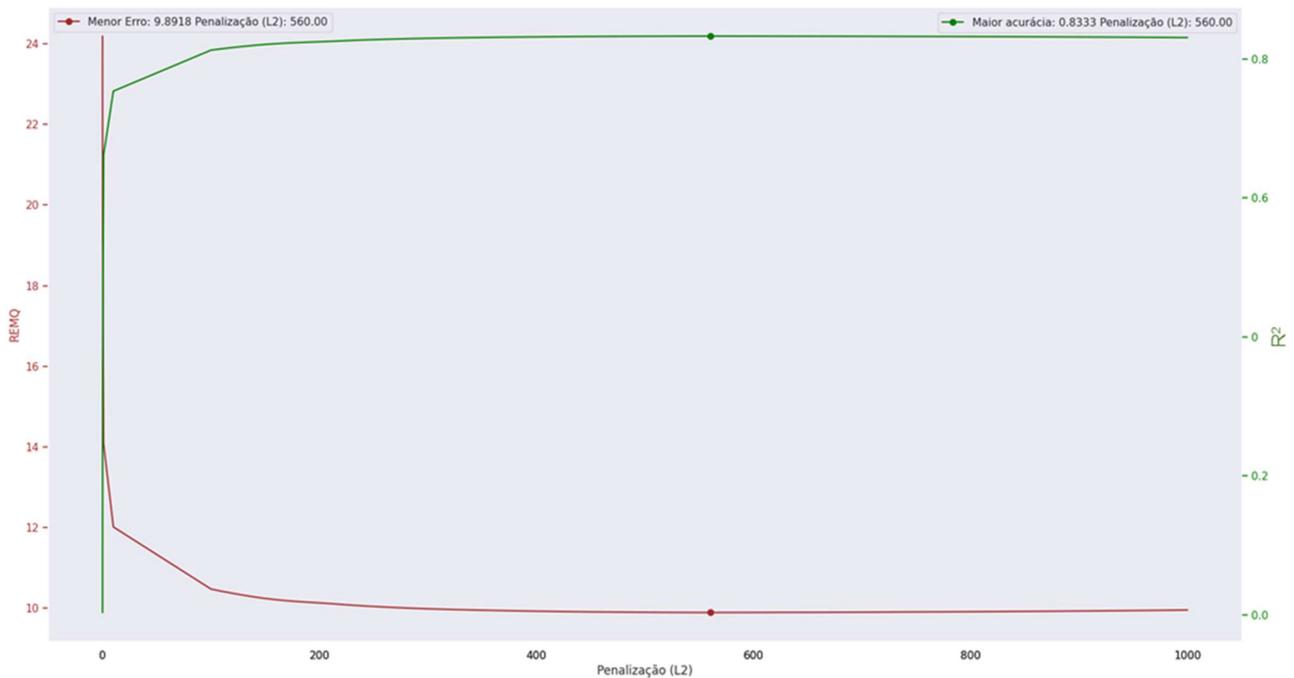


Figura 4.24 –  $RMSE$  e  $R^2$  do conjunto das rotas menores ou iguais a 600 km, para técnica de validação cruzada aplicada à otimização métricas das rotas dessa partição na busca pelo valor da penalização.

A aplicação do *Hold-out* na busca pela penalização nas rotas mais longas, para um modelo de SVR dedicado, retornou o valor de 10,0 (Figura 4.25) para o hiperparâmetro. Esse resultado levou a um  $R^2$  de 0,8814 e RMSE de 22,3488, valores estes mais próximos do esperado, com risco reduzido de sobre-ajuste. Apesar do  $R^2$  resultante ser levemente inferior àquela retornada pela abordagem de SVR única no contexto das rotas mais longas (0,8839), o uso da SVM dedicada trouxe a flexibilidade necessária para evitar o sobre-ajuste e o sub-ajuste, que estão sempre presentes no uso do modelo único, na partição cujas métricas não participaram dos processos de validação cruzada.

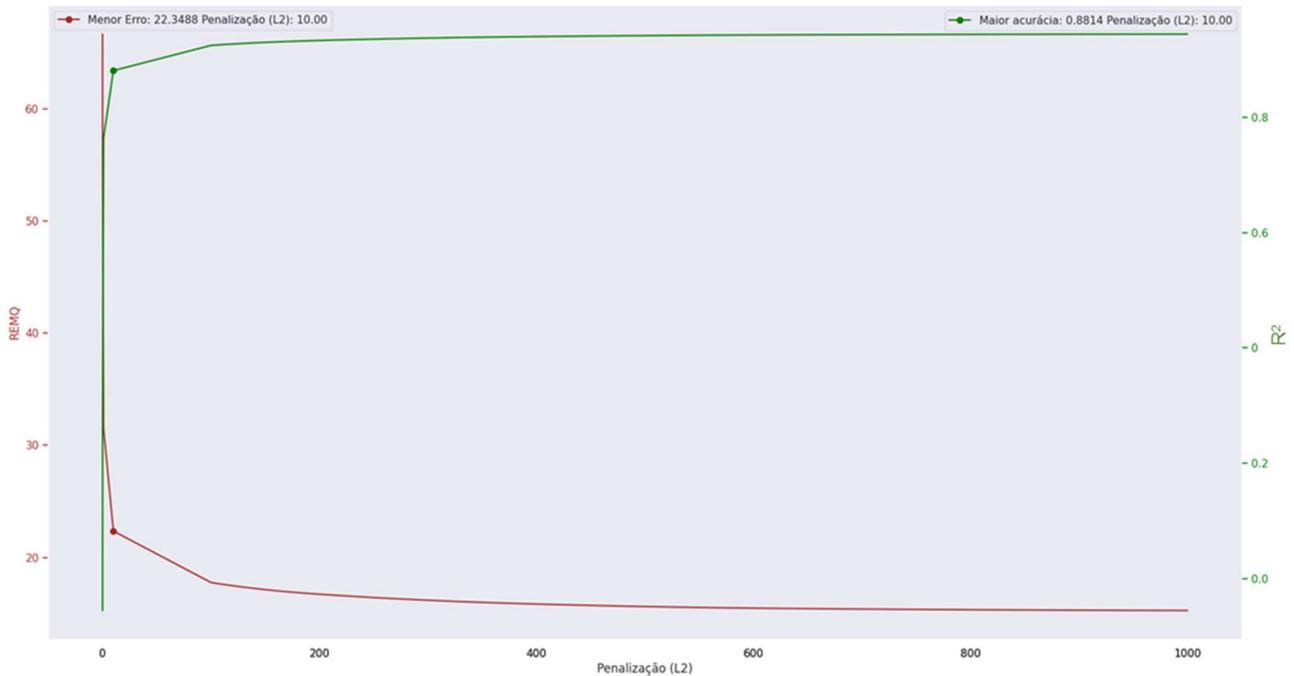


Figura 4.25 –  $RMSE$  e  $R^2$  do conjunto das rotas maiores que 600 km, para técnica de validação cruzada que otimiza métricas das rotas dessa partição na busca pelo valor da penalização.

A aplicação parametrização dos modelos individuais de SVR, com os valores identificados pelos respectivos processos de validação cruzada descritos nessa seção, resultou em um conjunto de valores melhor que o obtido pela SVR única, conforme vemos na Tabela 4.12. Para as rotas mais curtas a abordagem individual retornou um  $R^2$  de 0,8921, enquanto que o SVR único obteve 0,8470. Já na comparação com as rotas longas, a abordagem individual resultou em  $R^2$  de 0,8924 contra 0,8937 da SVR única, resultado este em favor da abordagem única. Todavia, a adoção da SVR única de rotas longas, traria um  $R^2$  de 0,7781 para as rotas curtas, valor inferior ao  $R^2$  alvo. Sendo assim a flexibilidade do uso de dois modelos individuais apresenta novamente a vantagem de se especializar modelos que respondam melhor (prevendo com mais exatidão) as particularidades de cada partição dos dados.

Tabela 4.12 – Resultados dos dois modelos de SVR independentes, aplicados sobre os conjuntos de treinamento e teste.

Otimização	$C$	$\Gamma$	$R^2$	RMSE
SVR rotas curtas	560,0	0,01	0,8921	8,0464
SVR rotas longas	10,0	0,01	0,8924	21,1167

Por último é importante apresentar o resultado dessa abordagem para o intervalo de rotas entre 600km e 900km, pois este foi um recorte em que a abordagem da MLP individual teve uma baixa performance. Inicialmente, quando analisamos o gráfico de dispersão da Figura 4.26, em que naquele intervalo os valores previstos pelo modelo (pontos vermelhos) quase que sobrepõe completamente os preços reais dos fretes. Existe uma diferença evidente quando comparamos com o gráfico da MLP individual da Figura 4.15, em que uma área relativamente grande de valores reais dos dados está aparente, sinalizando um alto erro. Essa intuição se confirma quando olhamos para o valor das métricas, em que a abordagem da SVR individual levou a um  $R^2$  de 0,7120 e RMSE de 14,1123, enquanto que o MLP análogo obteve 0,4869 e 18,8371 para as mesmas métricas.

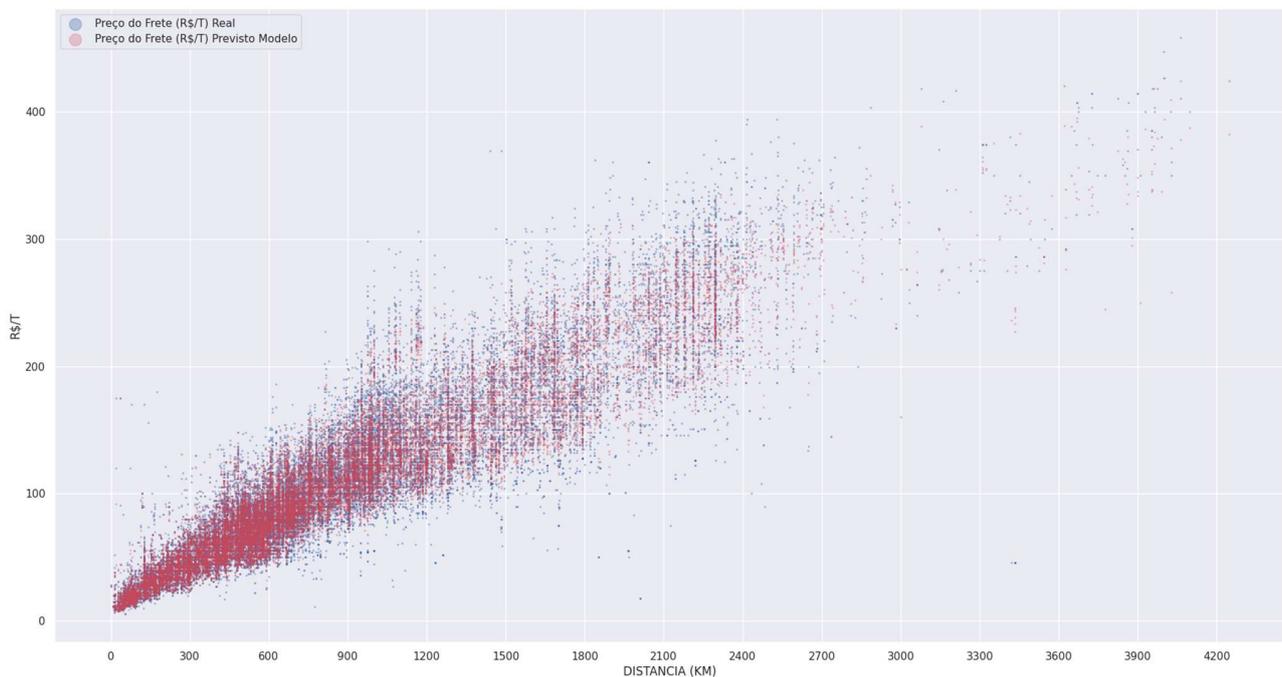


Figura 4.26 – Dispersão dos valores previstos pelos modelos individuais para rotas curtas e longas, e dos valores reais, com penalizações iguais a 560,0 e 10,0 respectivamente

### 4.3 MLP vs SVR

As técnicas de aprendizado de máquinas têm sido largamente utilizadas para previsões em uma variada gama de assuntos, tais como Biologia, Meteorologia, Medicina, planejamento de infraestrutura, inclusive a Logística. Na presente pesquisa, nossos resultados indicam que para a predição de fretes rodoviários de grãos agrícolas o melhor desempenho foi a técnica SVR que conseguiu alcançar melhores métricas de  $R^2$  e RMSE (Tabela 4.13) em detrimento à técnica MLP.

Outra verificação foi um melhor resultado obtido a partir da utilização de dois modelos individuais, que foram especializados em uma ou outra partição do conjunto das rotas até 600km (rotas curtas) e superiores à 600km (rotas longas), em detrimento ao uso de um modelo geral. Os valores previstos sobrepuseram-se de maneira bem distribuída em relação aos valores reais (Figura 4.26), mesmo após a aplicação da penalização não houve redução do  $R^2$  e tampouco aumento do erro, de maneira desproporcional a nenhum intervalo de distância em específico. As métricas mantiveram-se dentro dos intervalos aceitáveis, de modo a sinalizar um modelo com alto índice de acerto, sem que incorra no sobreajuste.

A partir da divisão dos dados em duas partições houve a necessidade de se especializar os modelos em um ou outro dos subconjuntos recém-criados, de forma a comparar seus resultados. Assim concebemos duas abordagens que foram aplicadas a cada uma das técnicas adotadas: a primeira em que um modelo é treinado sobre a totalidade dos dados, mas a otimização ocorre exclusivamente sobre as métricas de uma das partições; e uma segunda, em que treinamento e otimização se dão sobre a mesma partição, resultando em dois modelos. Na primeira abordagem apenas um dos modelos é escolhido para prever valores de frete, enquanto que na segunda ambos os modelos operam juntos, sendo acionado aquele que é especializado na extensão da rota ao qual pertença o valor de frete que se quer prever. Tomemos como exemplo a técnica MLP, cujos resultados são apresentados nas linhas 1 a 4 da Tabela 4.13. Nas linhas 1 e 2 podemos visualizar as métricas para os modelos da primeira abordagem, para rotas curtas e longas respectivamente. É possível identificar que quando a construção do modelo tem como foco a otimização de um subconjunto dos dados, o restante sofre com sobreajuste (linha 1) ou subajuste (linha 2). Por outro lado, modelos individuais que trabalham em conjunto (treinamento e otimização) sobre somente uma das partições apresentam resultados melhores (linhas 3 e 4).

A mesma lógica se aplica aos resultados obtidos para a técnica SVR, cujas métricas resultantes dos modelos otimizados sobre uma partição, ou outra, estão presentes nas linhas 5 e 7 (rotas curtas e

longas respectivamente) da Tabela 4.13, enquanto que os resultados para os modelos que operam em conjunto estão descritos nas linhas 7 e 8.

Tabela 4.13 – Resultados obtidos de SVR e MLP a partir do conjunto de teste

#Id	Técnica	Hiperparâmetros	$R^2 \leq 600$	RMSE $\leq 600$	$R^2 > 600$	RMSE $> 600$
1	MLP	Neurônios: 91 Penalização: 0,1	<b>0,8418</b>	<b>9,7448</b>	0,9473	14,9264
2	MLP	Neurônios: 91 Penalização: 100,0	0,5093	17,1649	<b>0,8279</b>	<b>26,9831</b>
3	MLP	Neurônios: 84 Penalização: 0,1	<b>0,8785</b>	<b>8,5404</b>		
4	MLP	Neurônios: 100 Penalização: 100,0			<b>0,8506</b>	<b>24,8818</b>
5	SVR	Gamma: 0,01 Penalização: 120,0	<b>0,8470</b>	<b>9,5854</b>	0,9353	16,5384
6	SVR	Gamma: 0,01 Penalização: 10,0	0,7781	11,5413	<b>0,8937</b>	<b>21,2037</b>
7	SVR	<b>Gamma: 0,01</b> <b>Penalização: 560,0</b>	<b>0,8921</b>	<b>8,0464</b>		
8	SVR	<b>Gamma: 0,01</b> <b>Penalização: 10,0</b>			<b>0,8924</b>	<b>21,1167</b>

Ince e Trafalis (2008), fizeram uso de Redes Neurais e Regressão de Vetor Suporte (a variante de Máquinas de Vetor Suporte, que executa regressão) na previsão do preço de ações, tendo como premissa que o valor futuro depende de indicadores financeiros, muito embora não haja um modelo paramétrico para explicar a relação, e por isso a opção por modelos orientados aos dados e não paramétricos. E assim puderam perceber uma melhor performance por parte da Regressão de Vetor Suporte, em detrimento à Rede Neural.

Em uso análogo ao proposto por esse projeto, Bowersox, Calantone e Rodrigues (2003) fizeram uso da técnica de Rede Neural para prever despesas de logística global, em termos percentuais do produto interno bruto por país. E com sucesso puderam atualizar os valores das mesmas despesas obtidas anteriormente em Bowersox e Cantalone (1998), em que fora utilizada a técnica pela primeira vez para prever aqueles valores.

No contexto da logística e cadeias de abastecimento, Liu, et al. (2016) faz uso de um modelo *Grey Neural Networks* (GNN) melhorado para prever súbitas interrupções na cadeia de transporte. Quando comparado com o modelo *Grey* tradicional (GM (1, 1)), e com o modelo GM (1, 1) melhorado, o GNN

melhorado apresenta um  $R^2$  mais elevada que o primeiro, e um pouco mais elevada que o segundo, apresentando-se assim como uma alternativa viável para a previsão de quebras na cadeia de abastecimento.

A decisão por rotas e modais de transporte estão muito ligadas ao preço final do frete (LOPES et al., 2020). O estudo realizado em Buijs et al., (2020), para identificar o impacto de uma nova linha de metrô na região metropolitana da cidade de Amsterdam, mesmo que no âmbito do transporte de pessoas, apresentou bons resultados. Os autores fizeram uso da técnica de clusterização K-means para identificar as melhores variáveis para o conjunto de treinamento a ser submetida à rede neural, que por sua vez deveria classificar o *dataset* representado pelo usuário. Os autores chegaram a resultados muito próximos da fronteira do sobre-ajuste. Posteriormente, puderam identificar que isso se deu em função da introdução de rotas não realizadas, geradas unicamente para aumentar o conjunto dos dados. A remoção de tais rotas reduziu o sobre-ajuste e tornou os resultados ainda mais factíveis.

No campo das *commodities* agrícolas, especificamente na produção de soja, Abraham, et al. (2020), aplicou a técnica das Redes Neurais Artificiais e algoritmos de Regressão, sobre uma série temporal, para prever área de colheita, rendimento e produção. A Rede Neural apresentou os melhores resultados para predição da área colhida e produção, enquanto que a Regressão Linear Simples implicou em uma maior taxa de acerto para a variável rendimento. Ainda assim a Rede Neural foi considerada uma opção superior uma vez que para séries temporais curtas (nesse caso 50 anos), o modelo apresentou bons resultados refletindo a robustez do modelo.

Quanto à aplicação de SVR para problemas relacionados à logística, Mak e Yang (2007) utilizaram a técnica para o transporte marítimo. Nesse contexto, a taxa de tratamento e transferência de containers (*container throughput*) em um terminal portuário é uma informação essencial para decisões ligadas a investimentos em infraestrutura. Sabendo disso, Mak e Yang (2007) modificaram a técnica SVM, dando luz ao modelo Least Squares Support Vector Machines (LS-SVM) com objetivo de prever o *container throughput* mensal no porto de Hong Kong. O LS-SVM sofreu modificações e definido como *Approximate* LS-SVM (ALSSVM), que resultaram num reduzido tempo de treinamento e menor consumo de memória. A performance do ALSSVM foi colocada à prova quando comparado ao SVM original, o LS-SVM, além do algoritmo de Rede Neural Artificial baseada em gradiente descendente. No resultado da comparação entre as técnicas o ALSSVM obteve o menor MAE juntamente com o LS-SVM, todavia, o primeiro chegou ao resultado fazendo menor uso de recursos (memória) e tempo de

processamento. Por fim, na métrica RMSE o ALSSVM obteve performance levemente pior do que o LS-SVM, 0,0096 do primeiro contra 0,0088 do segundo.

A produção de alimentos na Índia é largamente dependente da colheita de alguns cereais, inclusive o arroz. A sustentabilidade e a produtividade do arroz nas crescentes áreas de cultivo por sua vez dependem de condições climáticas adequadas. As bruscas variações sazonais do clima trazem um efeito negativo sobre a produção, inclusive a quebra de safra. Por isso, com objetivo de dispor de uma melhor previsibilidade das colheitas de arroz em 27 distritos do estado de Maharashtra, na Índia, Gandhi, et al. (2016) empregou a técnica SVM na previsão da colheita, em função de variáveis climáticas de precipitação, temperatura mínima, temperatura média, temperatura máxima, evaporação/transpiração, além de variáveis de área plantada e produção. A análise deu-se sobre um conjunto de dados do ano de 1998 a 2002, e as previsões atingiram um  $R^2$  de 78,76%, e um RMSE de 0,39.

O desempenho das técnicas SVM e MLP foram avaliadas por Hesami et al. (2020). O sucesso na cultura de tecidos vegetais depende diretamente da escolha do melhor protocolo de embriogênese somática, que é considerado o primeiro e mais importante passo no bem-sucedido estudo da transformação genética. Todavia a dificuldade de se chegar ao protocolo ótimo esbarra nos fatores de custo, tempo e complexidade, demandados pelo processo. No intuito contornar os obstáculos, Hesami, et al. (2020) empregou as técnicas de aprendizado de máquina MLP e *Support Vector Regression* (SVR) na modelagem da embriogênese somática de crisântemos (*chrysanthemum*). A comparação dos resultados demonstrou uma maior capacidade de acerto da técnica SVR em comparação à MLP. Isso se deu através da análise das métricas de  $R^2$  (coeficiente de determinação), RMSE e MAE, cujos valores encontrados foram, na ordem: 0,824, 15,403, 2,012 pela MLP, e 0,928, 10,697, 1,871 pela SVR, na busca pelo valor da taxa de calogênese. Da mesma forma a SVR superou a MLP nos demais experimentos, em que a variável a ser otimizada foi a taxa de embriogênese, e o número de embriões.

## 5 CONCLUSÕES

Este trabalho se propôs a testar a eficiência de predição de modelos gerados a partir de duas técnicas de aprendizagem de máquina, MLP e SVR, de forma a verificar se as técnicas de predição, quando submetidas à um conjunto de dados suficientemente representativos, são capazes de gerar modelos robustos para prever o comportamento do preço de frete de grãos agrícolas sob uma margem de erro aceitável. Além disso, esta pesquisa contribui disponibilizando uma ferramenta capaz de antecipar oscilações no mercado de transporte, bem como disponibilizar mais informações para auxiliar na mediação de conflitos entre os agentes que atuam no setor.

O transporte rodoviário apresenta-se como um grande integrador das principais regiões produtoras de grãos com os portos exportadores e o mercado doméstico. Em muitos casos, o frete rodoviário acaba sendo o balizador dos fretes dos outros modais; ou seja, é um indicador do mercado de fretes, tanto para o ferroviário quanto para o hidroviário.

O impacto financeiro com o transporte sentido pelos agentes econômicos que atuam nas cadeias agrícolas poderia ser reduzido com informações contribuíssem para uma maior previsibilidade do comportamento dos preços de transporte. A destacada participação do modal rodoviário ao longo do processo logístico de distribuição dos grãos faz com que a previsão dos custos de transporte relacionados a esta etapa seja uma importante ferramenta de gestão e tomada de decisão para os gestores do agronegócio.

Ambas as técnicas de predição empregadas foram submetidas à um conjunto de dados suficientemente representativos e geraram modelos robustos para prever o comportamento do preço de frete de grãos agrícolas. Tanto o SVR quanto o MLP foram eficientes para estimar os fretes rodoviários em termos de  $R^2$  e RMSE. Na comparação das técnicas, a técnica SVR teve um desempenho superior uma vez que alcançou melhores métricas na previsão do preço do transporte em relação à técnica MLP.

Em termos de particionamento do conjunto de dados em função da distância percorrida na rota de transporte, a utilização de dois modelos individuais especializados na partição do conjunto das rotas, um para até 600km e outro para distâncias superiores à 600km, tem um melhor desempenho quando comparado a adoção de um modelo geral para todas as faixas de distâncias. Isso confirma a hipótese de que os preços dos fretes rodoviários apresentam comportamento distintos para diferentes faixas de distâncias.

No Brasil, os avanços do agronegócio estão em sincronia por alguns setores da economia, tais como a ciência e a tecnologia. Por outro lado, o setor logístico não tem mostrado o mesmo desenvolvimento com a persistência de alguns entraves, como à falta de infraestrutura adequada para o transporte da produção agrícola e à incapacidade de armazenar adequadamente a colheita nacional. O atual desafio para o Estado é assegurar a manutenção da qualidade da infraestrutura já instaladas e promover um ambiente capaz de atrair o investimento privado para um novo ciclo de concessões.

## 6 REFERÊNCIAS

- ABIDOYE, R. B.; CHAN, A. P. C.; ABIDOYE, F. A.; OSHODI, O. S. Predicting property price index using artificial intelligence techniques: evidence from Hong Kong. **International Journal of Housing Markets and Analysis**, v. 12, n. 6, p. 1072-1092, 2019.
- ABRAHAM, E. R. et al. Time Series Prediction with Artificial Neural Networks: an analysis using Brazilian Soybean Production. **Agriculture**, v. 10, n. 10, 475, 2020.
- ARAÚJO, M. D. P. S.; BANDEIRA, R. A. DE M.; CAMPOS, V. ANIA B. G. Custos e fretes praticados no transporte rodoviário de cargas: uma análise comparativa entre autônomos e empresas. **Journal of Transport Literature**, v. 8, n. 4, p. 187-226, 2014.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. Nova Iorque: Springer, 2006
- BARTHOLOMEU, D. B.; PÉRA, T. G.; CAIXETA FILHO, J. V.; SU, C. T. Evaluation of green transport corridors of Brazilian soybean exports to China. **Brazilian Journal of Operations & Production Management**, v. 16, n. 3, p. 398-412, 2019.
- BLOIS, H. D.; MARTINS, R. S. A model of integration among prediction tools: applied study to road freight transportation. **RAUSP Management Journal**, v. 52, n. 1, p. 3-14, 2017.
- BOSER, B. E., GUYON, I. M.; VAPNIK, V. N. A Training Algorithm for Optimal Margin Classifiers. In: HAUSSLER, D. (Ed.). **Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92)**. Pittsburgh: ACM Press, 1992, p. 144-152.
- BOWERSOX, D. J.; CALANTONE, R. J. Executive Insights: Global Logistics. **Journal of International Marketing**, v. 6, n. 4, p. 83-93, 1998.
- BOWERSOX, D. J.; CALANTONE, R. J.; RODRIGUES, A. M. Estimation of global logistics expenditures using neural networks. **Journal of Business Logistics**, v. 24, n. 2, p. 21-36, 2003.
- BUIJS, R.; KOCH, T.; DUGUNDJI, E. Using Neural Nets to Predict Transportation Mode Choice: an Amsterdam Case Study. **Procedia Computer Science**, v. 170, p. 115-122, 2020.
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 2, n. 2, p. 121-167, 1998.
- CANGUSSU, S. V.; ROSA, M. J. A.; RODRIGUES, A. M. Determinantes do preço do frete rodoviário para transporte de soja em grão em Mato Grosso. **Revista de Economia da UEG**. v. 9, n. 1, p. 78-94, 2013.
- CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, v. 7, p. 1247-1250, 2014.
- CHIROMA, H.; ABDULKAREEM, S.; HERAWAN, T. Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. **Applied Energy**, v. 142, p. 266-273, 2015.
- COPPE, L. M.; DINIZ, D. A.; CARVALHO, V. H.; PIMENTEL, L. M. C. F. Precificação de fretes rodoviários via modelos paramétricos. **Revista Produção Online**, v. 20, n. 4, p. 1214-1237, 2020.
- CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 20, p. 273-297, 1995.
- Confederação Nacional do Transporte - CNT. **Boletim Unificado Dezembro 2021**. 2022. Disponível em: <https://www.cnt.org.br/boletins>. Acessado em: 10 jan. 2022.

CORREA, V. H. C.; RAMOS, P. A precariedade do transporte rodoviário brasileiro para o escoamento da produção de soja do Centro-Oeste: Situação e perspectivas. **Revista de Economia e Sociologia Rural**, v.48, n.2, p.447-472, 2010.

CORRÊA JUNIOR, G.; CAIXETA FILHO, J.V. Principais determinantes do preço do frete rodoviário para o transporte de soja em grãos em diferentes estados brasileiros: uma análise econométrica. **Economia Aplicada**, v. 7, n. 1, p. 189-211, 2003.

COSTA, G. F. **Modelo de previsão aplicado ao mercado de transporte rodoviário do açúcar no Estado de São Paulo**. 2014. 193p. Tese (Doutorado em Economia), Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo, Piracicaba.

CYBENKO, G. Approximations by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, v. 2, p. 303-314, 1989.

de BOK, M.; WESSELING, B.; KIEL, J.; MIETE, O.; FRANCKE, J. A sensitivity analysis of freight transport forecasts for The Netherlands. **International Journal of Transport Economics**, v. 45, n. 4, p. 571-587, 2018.

Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA. **Macrologística da Agropecuária Brasileira: delimitação das Bacias Logísticas**. 2016. Campinas: Embrapa Territorial, 2016. Disponível em: <https://www.embrapa.br/macrologistica/estudos-logisticos>. Acessado em: 10 set. 2020.

FERREIRA, J. F. K. **Comportamento do frete rodoviário de soja: aplicação de um modelo de regressão linear múltipla**. 2016. 53p. Dissertação (Mestrado em Pesquisa Operacional), Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas, Limeira.

GANDHI, N. et al. Rice Crop Yield Prediction in India using Support Vector Machines. In: 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, Khon Kaen. **Anais do JCSSE**, 2016, p. 1-5. Disponível em: <https://ieeexplore.ieee.org/document/7748856>. Acesso em: 10 jan. 2022.

GANGULI, S.; DUNNMON, J. **Machine Learning for better models for predicting bond prices**. 2017. Disponível em: <https://arxiv.org/abs/1705.01142>. Acessado em: 10 nov. 2020.

GONÇALVES, D. N. S.; GONÇALVES, C. M.; ASSIS, T. F.; SILVA, M. A. Analysis of the Difference between the Euclidean Distance and the Actual Road Distance in Brazil. **Transportation Research Procedia**, v. 3, p. 876-885, 2014.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.

GUENTHER, N.; SCHONLAU, M. Support vector machines. **The Stata Journal**, v.16, n. 4, p. 917-937, 2016.

GÜLER, H. An empirical modelling framework for forecasting freight transportation. **Transport**, v. 29, n. 2, p. 185-194, 2014.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3ed. Waltham: Morgan Kaufmann; Elsevier, 2012.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2nd ed. Nova Iorque: Springer, 2009.

HAYKIN, S. **Neural Networks and Learning Machines**. 3rd ed. Nova Jersey: Prentice Hall, 2008.

HESAMI, M.; et al. Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: effect of plant growth

- regulators on somatic embryogenesis of chrysanthemum, as a case study. **Plant Methods**, v. 16, n. 112, p. 1-15, 2020.
- INCE, H.; TRAFALISB, T. B. Short term forecasting with support vector machines and application to stock price prediction. **International Journal of General Systems**, v. 37, n. 6, p. 677-687, 2008.
- JANIC, M. Modelling the full cost of an intermodal and road freight transport network. **Transportation Research Part D**, v. 12, n. 1, p. 33-44, 2007.
- JOSSE, J.; HUSSON, F. Selecting the number of components in principal component analysis using cross-validation approximations. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1869-1879, 2012.
- KAVUSSANOS, M. G.; ALIZADEH-M, A. H. Seasonality patterns in dry bulk shipping spot and time charter freight rates. **Transportation Research Part E: Logistics and Transportation Review**, v. 37, n. 6, p. 443-467, 2001.
- LIMA, L. M.; ELIAS, L. P.; CAIXETA-FILHO, J. V.; COLETI, J. C. Fertilizer freight rate disparity in Brazil: a regional approach. **International Food and Agribusiness Management Review**, v. 19, n. 4, p. 109-128, 2016
- LIN, H.-T.; LIN, C.-J. A study on sigmoid kernels for SVM and the training of non- PSD kernels by SMO-type methods. **Technical report**, 2003. Disponível em: <https://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf>. Acesso em: 01 jul. 2021.
- LIU, C. et al. An Improved Grey Neural Network model for Predicting Transportation Disruptions. **Expert Systems with Applications**, v. 45, p. 331-340, 2016.
- LO, C-S; WANG, C-M. Support vector machine for breast MR image classification. **Computers & Mathematics with Applications**, v. 64, n. 5, p. 1153-1162, 2012.
- LOPES, H. S.; LIMA, R. S.; LEAL, F. Simulation Project for Logistics of Brazilian Soybean Exportation. **International Journal of Simulation Modelling**, v. 19, n. 4, p. 571-582, 2020.
- McCANN, P. A proof of the relationship between optimal vehicle size, haulage length and the structure of distance-transport costs. **Transportation Research Part A: Policy and Practice**, v. 35, n. 8, p. 671-693, 2001.
- MALDONADO, S.; MERIGÓ, J.; MIRANDA, J. Redefining support vector machines with the ordered weighted average. **Knowledge-Based Systems**, n. 148, p. 41-46, 2018.
- MARTINS, R. S. Estudo da formação do frete rodoviário e potencial de conflitos em negociações em cadeias do agronegócio brasileiro. **Organizações Rurais & Agroindustriais**, v. 10, n. 1, p. 73-87, 2008.
- MAK, K. L.; YANG, D. H. Forecasting Hong Kong's Container Throughput with Approximate Least Squares Support Vector Machines. In: World Congress on Engineering (WCE), 2007, London. **Anais do WCE**, 2007, p. 1-6. Disponível em: [http://www.iaeng.org/publication/WCE2007/WCE2007\\_pp7-12.pdf](http://www.iaeng.org/publication/WCE2007/WCE2007_pp7-12.pdf). Acesso em: 10 jan. 2022.
- MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biology**, v. 5, n. 4, p. 115-133, 1943.
- OLIVEIRA, A. L. R. Impactos da sazonalidade na safra de soja. **Análises e indicadores do agronegócio**, v. 1, p. 1-5, 2006.

- OLIVEIRA, A. L. R.; YAMAKAMI, A. Modelo de programação quadrática para análise da movimentação logística e comercialização da soja brasileira. **Engenharia Agrícola**, v. 26, n. 2, p. 552-560, 2006.
- OLIVEIRA, A. L. R.; CICOLIN, L.; SANTOS, M. C. Estimativa do custo rodoviário da soja: uma análise da rota Sorriso-Santos. **Revista de Economia e Agronegócio**, v. 11, n. 2, p. 255-276, 2013.
- OLIVEIRA, A. L. R. A logística do agronegócio: para além do apagão logístico. In: Buainain, A. M.; ALVES, E.; SILVEIRA, J. M. F. J.; NAVARRO, Z. (Orgs.). **O mundo rural no Brasil do século 21: a formação de um novo padrão agrário e agrícola**. 1ed. Brasília: Embrapa, 2014, p. 337-370.
- OLIVEIRA, A. L. R.; FILASSI, M.; LOPES, B. F. R.; MARSOLA, K. B. Logistical transportation routes optimization for Brazilian soybean: an application of the origin-destination matrix. **Ciência Rural**, v. 51, n. 2, p. 1-13, 2021.
- OLIVEIRA, A. L. R.; MARSOLA, K. B.; MILANEZ, A. P.; LIEGÉ, R. F. S. Performance evaluation of agricultural commodity logistics from a sustainability perspective. **Case Studies on Transport Policy**, v. 10, n. 1, p. 674-685, 2022.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. **Modelling transport**. 2nd. ed. Chichester: John Wiley & Sons, 1995.
- PERÁ, T. G. et al. Análise dos impactos da Medida Provisória nº 832 de 2018 (Política de Preços Mínimos do Transporte Rodoviário de Carga) na logística do agronegócio brasileiro. **Logística do Agronegócio: Desafios e Oportunidades**, v.3, 2018, 37p. Disponível em: <https://esalqlog.esalq.usp.br/categoria/serie-logistica-do-agronegocio>. Acesso em: 10 nov.2021.
- PRECHELT, L. Automatic early stopping using cross validation: quantifying the criteria. **Neural Networks**, v.11, p. 761-767, 1998a.
- PRECHELT, L. Early Stopping-But When? In: ORR, G. R.; Müller K. R. (Eds.) **Neural Networks: Tricks of the Trade, Lecture Notes in Computer Science**, no. 1524. Berlin: Springer, pp. 55-69, 1998b.
- PRINCIPE, J. C.; EULIANO, N. R.; LEFEBVRE, W. C. **Neural and adaptive systems: fundamentals through simulations**. New York: John Wiley & Sons, 1999.
- ROULET, M. C.; CAIXETA-FILHO, J. V. A Multiple-Criteria analysis application for vertical coordination in the transportation of agricultural commodities in Brazil. **Journal of Food Products Marketing**, v. 22, p. 1-13, 2016.
- RUMELHART, D., HINTON, G.; WILLIAMS, R. Learning representations by back-propagating errors. **Nature**, v. 323, 533-536, 1986.
- SAAB, A. A.; PAULA, R. A. O mercado de fertilizantes no Brasil: diagnósticos e propostas de políticas. **Revista de Política Agrícola**, v. 17, n. 2, p. 5-24, 2008.
- SCHÖLKOPF, B.; SMOLA, A. J. **Learning with Kernels**. Cambridge: MIT Press, 2002.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Neuro COLT Technical Report NC-TR-98-030**, Royal Holloway College, University of London, UK, 1998.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v.14, p.199-222, 2004
- SOARES, M. G.; CAIXETA-FILHO, J. V. Caracterização do mercado de fretes rodoviários para produtos agrícolas. **Gestão & Produção**, v.4, n.2, p. 186-204, 1997.

TANG, X.C.; MA Z.W.; HU Q.S. A Real-Time Arrhythmia Heartbeats Classification Algorithm Using Parallel Delta Modulations and Rotated Linear-Kernel Support Vector Machines. **IEEE Transactions on Bio-medical Engineering**, v. 67, n. 4, p. 978-986, 2020.

USDA. **Commodities & Products**. 2022. Disponível em: <https://www.fas.usda.gov/commodities>. Acesso em: 10 jan. 2022.

VAPNIK, V. **Statistical Learning Theory**. Nova Iorque: John Wiley and Sons, 1998.

WANKE, P. F. Fatores de satisfação com o uso de autônomos no transporte rodoviário de cargas. **Production**, v. 22, n. 3, p. 584-595, 2012.

## APÊNDICE

Apêndice 1. Detalhamento do Bando de Dados para os modelos preditivos de fretes rodoviários.

Agrupamento	Dados Colhidos	Recorte Geográfico	Recorte Temporal	Unidade	Fonte
<i>I</i>	<i>Frete (Preço e distância)</i>	Por Rota (Município de Origem para Município de Destino)	Mensal de 2012 a 2019	R\$/t e km	SIFRECA
<i>II</i>	<i>Produção dos Grãos</i>				
	Área Plantada	Por Estado e Município	Mensal de 2012 a 2019	Hectares	IBGE
	Produção	Por Estado e Município	Mensal de 2012 a 2019	Toneladas	IBGE
	Rendimento	Por Estado e Município	Mensal de 2012 a 2019	kg/hectare	IBGE
<i>III</i>	<i>Armazenagem</i>				
	Capacidade Estática	Por Estado e Por Município	Mensal de 2012 a 2019	Mil Toneladas	CONAB
<i>IV</i>	<i>Preço dos Grãos</i>				
	Cotação Internacional		Mensal de 2012 a 2019	US\$/t	CONAB
	Cotação Nacional	Por Estado e BR	Mensal de 2012 a 2019	R\$/t; R\$/saca(60kg)	CONAB; CEPEA
<i>V</i>	<i>Preço dos Combustíveis</i>				
	Preço do Diesel	Por Município e Estado	Mensal de 2012 a 2019	R\$/l	ANP
	Preço do Diesel S10	Por Município e Estado	Mensal de 2013 a 2019	R\$/l	ANP
	Preço do Etanol	Por Município e Estado	Mensal de 2012 a 2019	R\$/l	ANP
	Preço da Gasolina	Por Município e Estado	Mensal de 2012 a 2019	R\$/l	ANP
<i>VI</i>	<i>Exportações</i>				
	Petróleo	BR	Mensal de 2000 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/m <sup>3</sup>	COMEX STAT
	Gás Natural	BR	Mensal de 2000 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/m <sup>3</sup>	COMEX STAT
	Etanol Hidratado	BR	Mensal de 2012 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/m <sup>3</sup>	COMEX STAT
	Etanol Anidro	BR	Mensal de 2012 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/m <sup>3</sup>	COMEX STAT
	Milho	Por Município e Porto	Mensal de 2012 a 2019	Toneladas; Valor FOB (US\$)	COMEX STAT
	Soja	Por Município e Porto	Mensal de 2012 a 2020	Toneladas; Valor FOB (US\$)	COMEX STAT
<i>VII</i>	<i>Importação</i>				
	Derivados de Petróleo	BR	Mensal de 2000 a 2020	m <sup>3</sup>	COMEX STAT
	Petróleo	BR	Mensal de 2000 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/m <sup>3</sup>	COMEX STAT
	Etanol Anidro	BR	Mensal de 2012 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/ m <sup>3</sup>	COMEX STAT
	Etanol Anidro	BR	Mensal de 2012 a 2020	m <sup>3</sup> ; Valor FOB(US\$)/ m <sup>3</sup>	COMEX STAT

(continua)

(continuação)

<b>Agrupamento</b>	<b>Dados Colhidos</b>	<b>Recorte Geográfico</b>	<b>Recorte Temporal</b>	<b>Unidade</b>	<b>Fonte</b>
<i>VIII</i>	<i>Índices</i>				
	IGPM	BR	Mensal e Anual de 2012 a 2019	%	FGV
	IPCA	BR	Mensal e Anual de 2012 a 2019	%	IBGE
<i>IX</i>	<i>Câmbio (Taxa de Câmbio)</i>	Dólar EUA	Mensal e Anual de 2012 a 2019	R\$	Banco Central
	Moagem	Por Município	Anual 2010 e 2014	Toneladas/dia	ABIOVE
	Capacidade Instalada de Óleos e Vegetais	Por Estado	Anual de 2012 a 2019	Toneladas/dia	ABIOVE
<i>XI</i>	<i>Safra</i>				
	Período de safra (colheita)	Por cultura	Safra e entressafra	Meses	CONAB