

UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

Rodrigo Lemos Godinho

**Análise comparativa entre técnicas de
aprendizado de máquina aplicadas para a
predição de preços de produtos hortifrutícolas**

São Carlos - SP

2023

Rodrigo Lemos Godinho

Análise comparativa entre técnicas de aprendizado de máquina aplicadas para a predição de preços de produtos hortifrutícolas

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia de Computação da Universidade Federal de São Carlos, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação.

Orientação Prof. Dr. Alan Demétrius Baria Valejo

São Carlos - SP
2023

Dedico este trabalho àqueles que me inspiram a ser uma pessoa melhor.

Agradecimentos

Agradeço aos meus pais, que me apoiaram nos momentos difíceis e me deram a chance de seguir meus sonhos.

Agradeço ao meu orientador, prof. Alan, pela oportunidade cedida, todo o conhecimento e o apoio na confecção deste trabalho.

Por fim, agradeço aos meus amigos, que estiveram juntos de mim em cada passo desta jornada.

Podemos ver apenas uma curta distância à frente, mas podemos ver que há muito lá a ser feito.
(Alan Turing)

Resumo

A agricultura familiar é caracterizada como toda forma de cultivo de terra administrada por uma família, empregando como principal mão de obra seus próprios membros. A maior parte dos estabelecimentos agrícolas do Brasil se enquadram nessa definição, contudo, a área ocupada por esses agricultores e os recursos disponíveis são muito inferiores em comparação aos grandes produtores. Com área e recursos diminutos, é de suma importância que os pequenos agricultores sejam capazes de otimizar sua produção para garantir a sustentabilidade do seu negócio. O objetivo deste trabalho é estudar e comparar o desempenho de diferentes técnicas de Aprendizado de Máquina na tarefa de predição de séries temporais do preço de produtos hortifrutícolas para períodos de curto prazo, a fim de auxiliar os pequenos agricultores na negociação de um preço justo para sua mercadoria. A partir dos resultados obtidos pelos modelos induzidos, foi possível concluir que as técnicas de Florestas Aleatórias, Regressão por Vetores Suporte e Redes de Memória Longa de Curto Prazo têm um desempenho parecido na tarefa de predição de preços de produtos hortifrutícolas. Além disso, os modelos apresentam potencial para serem utilizados como ferramentas capazes de auxiliar o produtor no momento de venda dos seus produtos.

Palavras-chave: Aprendizado de Máquina; Séries Temporais; Análise Preditiva; Hortifrutícolas.

Abstract

Family farming is characterized as any form of land cultivation managed by a family, employing its own members as the main labor force. Most agricultural establishments in Brazil fit this definition, however, the area occupied by these farmers and the available resources are much smaller compared with that of large agricultural companies. With limited area and resources, it is extremely important that small farmers are able to optimize their production to ensure the sustainability of their business. The objective of this work is to study and compare the performance of different machine learning techniques in the task of predicting time series of horticultural products prices for short term periods, in order to help small farmers to negotiate a fair price for their merchandise. From the results obtained by the induced models, it was possible to conclude that the Random Forests, Support Vector Regression and Long Short-Term Memory techniques have a similar performance in the task of predicting prices of fruits and vegetables. In addition, the models have the potential to be used as tools capable of assisting producers when selling their merchandise.

Keywords: Machine Learning; Time Series; Forecasting; Horticultural.

Lista de ilustrações

| | |
|---|----|
| Figura 1 – Visão geral de como aplicar AM na resolução de uma determinada tarefa. | 25 |
| Figura 2 – Gráfico da série temporal da população dos EUA. | 26 |
| Figura 3 – Diagrama simplificado do funcionamento de uma RF. | 28 |
| Figura 4 – Diagrama simplificado do funcionamento de uma SVM num espaço de 2 dimensões. | 28 |
| Figura 5 – Diagrama simplificado do algoritmo de SVR num espaço de 2 dimensões. | 29 |
| Figura 6 – Diagrama de uma rede neural. | 30 |
| Figura 7 – Diagrama de uma RNR. | 31 |
| Figura 8 – Diagrama de uma LSTM. | 31 |
| Figura 9 – Esquema das tabelas do banco de dados. | 36 |
| Figura 10 – Série temporal dos preços diários do Tomate. | 37 |
| Figura 11 – Esquema do conjunto de aprendizado supervisionado. | 38 |
| Figura 12 – Previsão de 7 dias utilizando RF. | 41 |
| Figura 13 – Previsão de 30 dias utilizando RF. | 42 |
| Figura 14 – Previsão de 90 dias utilizando RF. | 42 |
| Figura 15 – Previsão de 180 dias utilizando RF. | 42 |
| Figura 16 – Previsão de 7 dias utilizando SVR. | 43 |
| Figura 17 – Previsão de 30 dias utilizando SVR. | 43 |
| Figura 18 – Previsão de 90 dias utilizando SVR. | 43 |
| Figura 19 – Previsão de 180 dias utilizando SVR. | 44 |
| Figura 20 – Previsão de 7 dias utilizando LSTM. | 44 |
| Figura 21 – Previsão de 30 dias utilizando LSTM. | 44 |
| Figura 22 – Previsão de 90 dias utilizando LSTM. | 45 |
| Figura 23 – Previsão de 180 dias utilizando LSTM. | 45 |
| Figura 24 – Gráfico de box-plot para os valores do MAE de cada algoritmo de AM estudado. | 47 |
| Figura 25 – Gráfico de box-plot para os valores do MAPE de cada algoritmo de AM estudado. | 48 |
| Figura 26 – Gráfico de box-plot para os valores do MSE de cada algoritmo de AM estudado. | 49 |
| Figura 27 – Gráfico de box-plot para os valores do RMSE de cada algoritmo de AM estudado. | 49 |
| Figura 28 – Gráfico dos resultados da simulação com o produto Tomate. | 50 |

Lista de tabelas

| | |
|--|----|
| Tabela 1 – Trabalhos relacionados. | 33 |
| Tabela 2 – Métricas de performance para a previsão de 7 dias do Tomate. | 46 |
| Tabela 3 – Métricas de performance para a previsão de 30 dias do Tomate. | 46 |
| Tabela 4 – Métricas de performance para a previsão de 90 dias do Tomate. | 46 |
| Tabela 5 – Métricas de performance para a previsão de 180 dias do Tomate. | 47 |
| Tabela 6 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando RF. | 51 |
| Tabela 7 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando SVR. | 51 |
| Tabela 8 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando LSTM. | 51 |
| Tabela 9 – Métricas de performance das previsões do produto Alface | 60 |
| Tabela 10 – Métricas de performance das previsões do produto Banana | 61 |
| Tabela 11 – Métricas de performance das previsões do produto Batata | 62 |
| Tabela 12 – Métricas de performance das previsões do produto Cebola | 63 |
| Tabela 13 – Métricas de performance das previsões do produto Cenoura | 64 |
| Tabela 14 – Métricas de performance das previsões do produto Laranja | 65 |
| Tabela 15 – Métricas de performance das previsões do produto Maçã | 66 |
| Tabela 16 – Métricas de performance das previsões do produto Mamão | 67 |
| Tabela 17 – Métricas de performance das previsões do produto Melancia | 68 |
| Tabela 18 – Métricas de performance das previsões do produto Tomate | 69 |

Sumário

| | | |
|------------|--|-----------|
| | Lista de ilustrações | 13 |
| | Lista de tabelas | 15 |
| 1 | INTRODUÇÃO | 19 |
| 1.1 | Justificativas | 19 |
| 1.2 | Objetivos | 20 |
| 1.2.1 | Objetivo Geral | 20 |
| 1.2.2 | Objetivos Específicos | 20 |
| 1.3 | Organização do Trabalho | 21 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 23 |
| 2.1 | Inteligência Artificial | 23 |
| 2.2 | Aprendizado de Máquina | 24 |
| 2.3 | Séries Temporais | 25 |
| 2.3.1 | Florestas Aleatórias | 27 |
| 2.3.2 | Regressão por Vetores Suporte | 27 |
| 2.3.3 | Redes de Memória Longa de Curto Prazo | 30 |
| 2.3.4 | Métricas de performance | 32 |
| 2.4 | Trabalhos relacionados | 32 |
| 3 | METODOLOGIA | 35 |
| 3.1 | Conjunto de dados | 35 |
| 3.2 | Aquisição e preparação dos dados | 36 |
| 3.3 | Mineração dos dados pelos modelos de AM | 37 |
| 3.3.1 | Parâmetros de configuração das técnicas de AM | 38 |
| 4 | ANÁLISE EXPERIMENTAL E RESULTADOS | 41 |
| 4.1 | Previsões geradas | 41 |
| 4.2 | Avaliação de desempenho preditivo dos modelos de AM | 44 |
| 4.2.1 | Avaliação a partir de métricas estatísticas | 45 |
| 4.2.2 | Simulação de desempenho financeiro | 50 |
| 5 | CONCLUSÃO | 53 |
| | REFERÊNCIAS | 55 |

| | | |
|----------|----------------------------|-----------|
| A | APÊNDICES | 59 |
|----------|----------------------------|-----------|

1 Introdução

Na década de 1960, observa-se em todo o território nacional uma desorganização estrutural em todas as etapas e processos responsáveis pelo sistema de produção e distribuição de produtos hortifrutícolas. A situação dificultava o abastecimento de alimentos agrícolas nas grandes concentrações populacionais do país (BARRETO JÚNIOR, 2013).

Com o objetivo de melhorar a qualidade da comercialização dessa classe de produtos ao redor do país, o Governo Federal criou as Centrais de Abastecimento (CEASAs). As CEASAs são empresas estatais ou de capital misto destinadas a promover, desenvolver, regular, dinamizar e organizar a comercialização de produtos da hortifruticultura a nível de atacado em uma região. Qualquer produtor pode comercializar livremente seus produtos em uma Central pelo preço que quiser, sejam eles provenientes de agricultura patronal ou familiar (ABRACEN, 2007; CEASA/CE, 2016).

Agricultura familiar é o nome atribuído a toda forma de cultivo de terra administrada por uma família, empregando como principal mão de obra os membros da mesma. Ela é caracterizada por ocorrer em pequenas propriedades com alta diversidade produtiva, que muitas vezes alia a produção de subsistência a uma produção destinada ao mercado (BRASIL, 2006; Embrapa, 2021).

O Censo Agropecuário de 2017 aponta que 77% dos estabelecimentos agrícolas do país foram classificados como de agricultura familiar. Em extensão de área, a agricultura familiar ocupava no período da pesquisa 80,9 milhões de hectares, o que representa 23% da área total dos estabelecimentos agropecuários brasileiros (IBGE, 2019).

Segundo o Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE), no ano de 2020, do total de 3.897.408 de estabelecimentos da agricultura familiar no Brasil, 3.167.286 possuíam uma produção mensal de até 3 mil reais (DIEESE, 2020).

Para o pequeno produtor, com poucos recursos, as CEASAs são um grande facilitador para os seus negócios, pois fornecem um espaço para a armazenagem de produtos, agilizam o contato com varejistas e até mesmo auxiliam na tarefa de precificação, através da divulgação de boletins diários com os preços médios vigentes dos produtos comercializados (CEASA/CE, 2016).

1.1 Justificativas

Apesar de serem a maior parcela de agricultores no país segundo o IBGE (2019), os agricultores familiares ainda enfrentam muitas desvantagens.

Dos mais de 5 milhões de estabelecimentos de agricultura familiar, apenas 1,4 milhões de produtores declararam ter acesso à internet. Cerca de 15,5% dos produtores disseram nunca ter frequentado escola e 73% frequentaram somente o ensino fundamental. Além disso, 23,03% dos produtores declararam não saber ler e escrever (IBGE, 2019).

Se tratando de pequenos agricultores, com recursos extremamente limitados, é imprescindível que esses recursos sejam alocados de maneira ótima, a fim de garantir a manutenção de suas atividades.

O prosseguimento deste trabalho se justifica não apenas pelo levantamento de resultados relevante sobre o estudo de técnicas de Aprendizado de Máquina para a predição de séries temporais, mas principalmente para ajudar os pequenos agricultores no processo de gestão de mercadoria, de modo que sejam capazes de identificar os melhores dias para comercializar seus produtos e maximizar seus retornos.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral do presente trabalho é comparar o desempenho de diferentes técnicas de AM na tarefa de predição do preço de produtos hortifrutícolas comercializados na Companhia de Entrepósitos e Armazéns Gerais de São Paulo (CEAGESP).

1.2.2 Objetivos Específicos

Os objetivos específicos podem ser desmembrados nos seguintes:

- Coletar os dados de preços dos produtos para um período suficiente para realizar o treinamento dos modelos de Aprendizado de Máquina;
- Realizar um pré-processamento nos dados para adequá-los a cada um dos modelos;
- Treinar os modelos e realizar as previsões de preços para diferentes períodos de tempo;
- Avaliar e comparar o desempenho das técnicas de Aprendizado de Máquina utilizando medidas estatísticas de performance;
- Avaliar e comparar o desempenho das técnicas de Aprendizado de Máquina considerando a simulação de um cenário em que um produtor utiliza os modelos gerados para escolher o momento ideal para a venda de seus produtos.

1.3 Organização do Trabalho

O trabalho em questão está organizado da seguinte forma:

- No Capítulo 2 é apresentada a fundamentação teórica do trabalho. Os estudos que serviram como fundação deste trabalho e direcionaram o seu desenvolvimento foram apresentados e utilizados para definir cada um dos conceitos abordados ao longo dos experimentos.
- No Capítulo 3 são apresentados os experimentos realizados. Toda a metodologia utilizada na coleta e preparação dos dados e no treinamento dos modelos é discutida e detalhada;
- No Capítulo 4 são apresentados os resultados obtidos das predições realizadas para cada um dos modelos de AM e as comparações feitas entre cada um deles, além da discussão acerca do porque tais resultados foram obtidos;
- No Capítulo 5, por fim, são apresentadas as conclusões obtidas ao longo do desenvolvimento deste trabalho.

2 Fundamentação Teórica

Neste capítulo são abordados os aspectos específicos das teorias utilizadas para o desenvolvimento deste trabalho, apresentando juntamente os estudos que serviram de embasamento para os experimentos e as análises subsequentes. São delineados os conceitos básicos de Inteligência Artificial, Aprendizado de Máquina e Séries Temporais. Por fim, são discutidos a tarefa de Previsão de Séries Temporais e os princípios gerais do funcionamento de cada uma das técnicas de Aprendizado de Máquina implementadas para a resolução do problema proposto.

2.1 Inteligência Artificial

Inteligência Artificial (IA) é um campo de estudo que abrange várias áreas do conhecimento de ciência da computação. O termo recebeu diversas definições ao longo da história. McCarthy (2004) definiu IA como a ciência e engenharia de construção de máquinas inteligentes, especialmente programas de computador.

Uma das dificuldades para se encontrar uma definição capaz de encapsular todas as dimensões da IA se deve à própria definição de inteligência. Turing (1950), na tentativa de definir quais requisitos seriam necessários para julgar que uma máquina possuía inteligência, propôs o Teste de Turing. Para ser definida como “inteligente”, uma máquina deveria satisfazer as seguintes condições:

- **Processamento de linguagem natural:** ser capaz de se comunicar em linguagem humana;
- **Representação de conhecimento:** armazenar conhecimento;
- **Raciocínio automático:** utilizar a informação armazenada para responder perguntas e gerar novas conclusões;
- **Aprendizado de máquina:** adaptar a novas circunstâncias e detectar e extrapolar padrões.

Para realizar as predições deste trabalho, foi feito o uso de IA, através de técnicas que aplicam o conceito de Aprendizado de Máquina.

2.2 Aprendizado de Máquina

Aprendizado de Máquina (AM), um subcampo da IA, é o estudo sistemático de algoritmos que expandem seu conhecimento ou melhoram sua performance através de experiência obtida via de dados de treinamento, sem programação explícita. O objetivo final do AM é gerar uma aplicação capaz de resolver uma tarefa para a qual foi treinada (FLACH, 2012).

Para extrair um conhecimento desejado de um certo domínio ou conjunto de dados, o primeiro passo é identificar seus atributos e variáveis capazes de descrever o domínio em questão. A partir dos atributos, é preciso escolher uma tarefa a ser resolvida. No contexto de AM, algumas tarefas comumente realizadas são:

- **Classificação:** dividir os dados de entrada em duas ou mais classes. Essas classes são um conjunto discreto de valores conhecido previamente;
- **Regressão:** deduzir uma função que relaciona os dados de entrada a um valor contínuo;
- **Agrupamento:** dividir os dados de entrada em grupos. Diferentemente da tarefa de classificação, os grupos não são conhecidos previamente e, nesse caso, o objetivo é encontrar estruturas naturais nos dados.

Para realizar uma tarefa, é preciso utilizar um algoritmo de AM. Esse algoritmo é alimentado com um conjunto de dados de treinamento. Esse conjunto faz parte do domínio de interesse e apresenta todos os atributos identificados inicialmente. A partir desses dados, é gerado um modelo. O modelo é o produto da tentativa do algoritmo de aprendizado de encontrar padrões nos atributos dos dados de treinamento para determinar um resultado. Se um modelo leva em consideração somente um atributo, ele pode ser chamado de modelo univariado. Caso contrário, ele é um modelo multivariado (FLACH, 2012).

Os modelos podem, também, ser divididos de acordo com a natureza do seu aprendizado, como, por exemplo, Aprendizado Supervisionado e Aprendizado Não-Supervisionado. No caso do Aprendizado Supervisionado, o algoritmo é alimentado com pares de entradas e saídas desejadas e tem como objetivo encontrar uma função que mapeia valores de novas entradas para os valores corretos de saída. Tarefas comuns desse tipo de Aprendizado são Classificação e Regressão. Já no caso do Aprendizado Não-Supervisionado, o algoritmo fica encarregado de identificar características naturais nos dados de entrada para gerar um novo resultado. Uma tarefa comum de Aprendizado Não-Supervisionado é a de Agrupamento. Ressalta-se que existem outros tipos de de aprendizado, como, por exemplo, Aprendizado Semisupervisionado e Aprendizado por Reforço (FLACH, 2012).

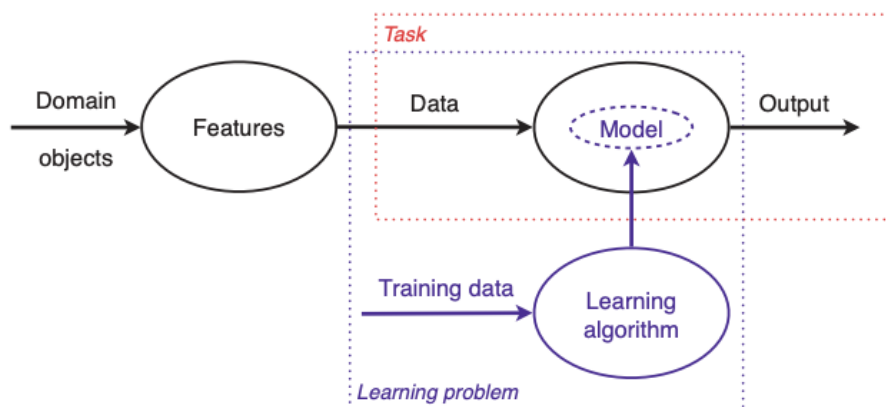
Quando o resultado de um modelo é a tentativa de prever um valor a partir de suas variáveis de entrada, ele pode ser chamado de modelo preditivo. Já quando o objetivo do

modelo está relacionado à descrição dos dados, ele pode ser chamado de modelo descritivo (FLACH, 2012).

Com o modelo obtido através dos dados de treinamento, é possível avaliar sua performance em cima de novos dados e determinar se o resultado é satisfatório para a resolução da tarefa proposta. A performance de um modelo em cima dos dados de treinamento não é relevante, pois o modelo precisa ser capaz de generalizar o conhecimento obtido e gerar novas conclusões a partir de dados nunca antes vistos. Um foco exagerado no desempenho sobre os dados de treinamento pode fazer com que o modelo dê uma ênfase indevida a padrões e características encontrados somente naquele conjunto de amostras. A performance do modelo melhora melhora durante o treinamento, porém ele está tão ajustado para aqueles dados específicos que ocorre uma queda de performance na tentativa de aplicar o conhecimento adquirido a dados novos. Esse fenômeno é conhecido como sobreajuste, do inglês *overfitting* (FLACH, 2012).

A Figura 1 ilustra a aplicação de uma técnica de AM na resolução de uma determinada tarefa.

Figura 1 – Visão geral de como aplicar AM na resolução de uma determinada tarefa.



Fonte: (FLACH, 2012).

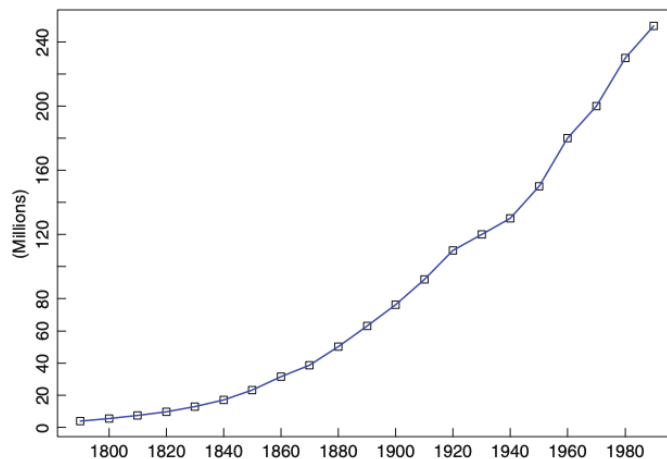
Neste trabalho, as técnicas de AM escolhidas farão uso de um conjunto univariado de dados para treinar modelos preditivos (tarefa de Regressão) de aprendizado supervisionado.

2.3 Séries Temporais

Séries temporais são conjuntos de observações x_t , cada uma registrada em um instante de tempo t . Elas podem ser classificadas como discretas ou contínuas. Uma série temporal discreta, como as séries temporais analisadas neste trabalho, é uma série em que o conjunto T_0 de instantes em que uma observação é registrada é um conjunto discreto, por exemplo, se

as observações são registradas em intervalos de tempo fixos. A Figura 2 ilustra um exemplo de um gráfico de uma série temporal discreta, na qual foram feitas observações em intervalos de 10 anos sobre a população total dos Estados Unidos.

Figura 2 – Gráfico da série temporal da população dos EUA.



Fonte: (BROCKWELL; DAVIS, 2016).

A análise de séries temporais é um campo da estatística que estuda dados amostrados ao longo do tempo para criar e validar hipóteses sobre os mesmos. Uma tarefa comum da análise de séries temporais é a Previsão, do inglês *Forecasting*. O objetivo desta tarefa é prever o futuro a partir de um conjunto de informações disponíveis, incluindo dados históricos e conhecimento sobre quaisquer eventos futuros que possam impactar a previsão, por meio de um modelo. Este modelo pode ser uma técnica estatística ou de AM capaz de extrair características e padrões dos dados analisados, para que seja possível supor qual será o comportamento da série no futuro (BROCKWELL; DAVIS, 2016; HYNDMAN; ATHANASOPOULOS, 2021).

Segundo Hyndman e Athanasopoulos (2021), a tarefa de *forecasting* pode ser decomposta em 5 passos:

1. Definição do problema a ser analisado;
2. Coleta dos dados históricos que serão utilizados para treinar os modelos;
3. Análise e pré-processamento dos dados;
4. Treinamento dos modelos preditivos;
5. Uso e avaliação dos modelos.

Neste trabalho, foram selecionados três técnicas de AM comumente utilizadas na literatura na tarefa de predição de preços: Florestas Aleatórias (KHAIDEM; SAHA; DEY, 2016),

Regressão por Vetores Suporte (TAY; CAO, 2001) e Redes de Memória Longa de Curto Prazo (NASSAR et al., 2020).

2.3.1 Florestas Aleatórias

As Florestas Aleatórias, do inglês *Random Forest* (RF), são um método de AM supervisionado que pode ser utilizado tanto para tarefas de regressão, quanto de classificação (TIBCO, 2021b).

De maneira simplificada, RF são um conjunto de árvores de decisão. Cada árvore é composta por vários nós, cada nó representando uma característica diferente do conjunto de dados. O algoritmo tenta agrupar os valores de entrada através de sucessivas classificações de acordo com as características que compõem a árvore. Nas RF, os atributos que compõem cada árvore de decisão são definidos a partir de um vetor aleatório e amostrado de forma independente. Segundo a literatura, a seleção aleatória de atributos contribui para a geração de árvores com baixa correlação, de forma a minimizar o erro individual de cada uma (FLACH, 2012; BREIMAN, 2001).

Depois de treinado, o modelo possui um conjunto de árvores de decisão em seu interior. Quando o problema a ser solucionado se trata de uma classificação, a predição da floresta é determinada pela maioria de votos das árvores que compõem a floresta. Se o problema em questão for uma regressão, a média dos resultados encontrados pelas árvores é utilizada (TIBCO, 2021b).

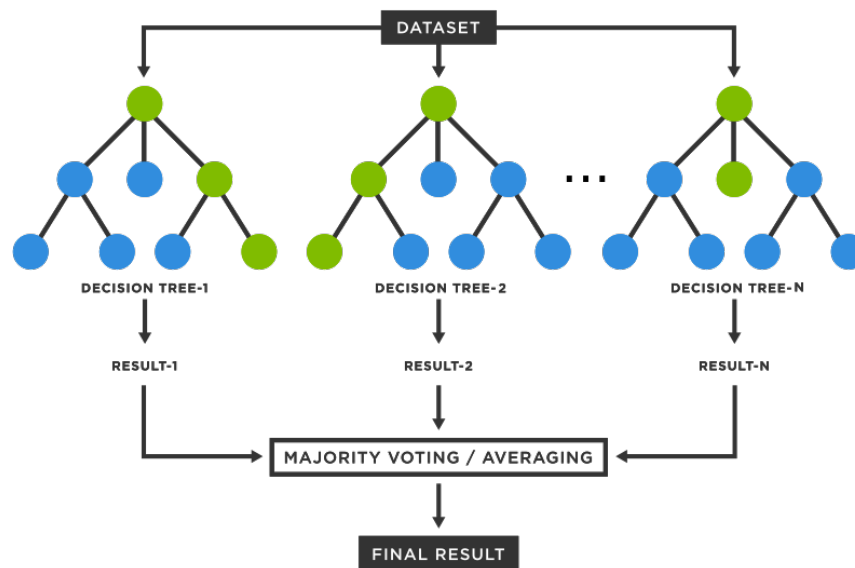
A Figura 3 ilustra o funcionamento de uma RF, em que um conjunto de árvores de decisão combinam seus resultados para gerar uma resposta para o algoritmo.

2.3.2 Regressão por Vetores Suporte

Máquina de vetores de suporte, do inglês *Support Vector Machine* (SVM), é um método de AM supervisionado utilizado em tarefas de classificação. O algoritmo funciona de forma a encontrar um hiperplano num espaço de n dimensões que melhor divide os dados em diferentes classes. Os pontos que estão mais próximos do hiperplano são chamados de vetores de suporte. A distância entre o hiperplano e os vetores de suporte é chamada de margem. Portanto, o hiperplano que melhor divide os dados é aquele que maximiza o tamanho da margem (FLACH, 2012; MATHWORKS, 2016).

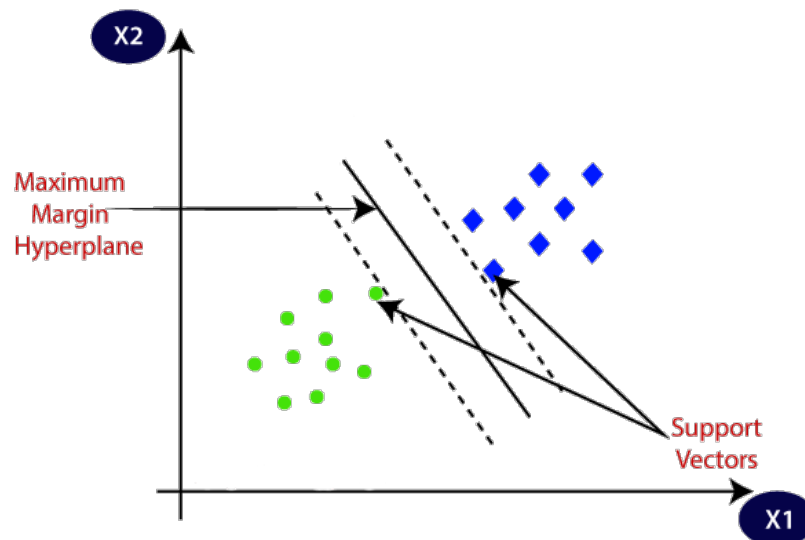
A Figura 4 ilustra um hiperplano, suas margens e os vetores de suporte de uma SVM treinada que dividiu seus dados de entrada em duas classes diferentes.

Figura 3 – Diagrama simplificado do funcionamento de uma RF.



Fonte: Adaptado de (TIBCO, 2021b).

Figura 4 – Diagrama simplificado do funcionamento de uma SVM num espaço de 2 dimensões.



Fonte: Adaptado de (JAVATPOINT, 2019).

Drucker et al. (1996) propôs uma versão de SVM para tratar de tarefas de regressão, chamada de Regressão por Vetores Suporte, do inglês *Support Vector Regression* (SVR). O algoritmo de SVR busca um hiperplano que contém o maior número de pontos possível dos dados dentro de uma margem de erro definida (DRUCKER et al., 1996; MATHWORKS, 2015).

Para exemplificar o funcionamento do algoritmo de SVR, de forma simplificada, o hiperplano encontrado em um espaço de duas dimensões pode ser definido como uma reta descrita pela Equação 2.1, em que w_i é o coeficiente que multiplica a variável atributo x_i para encontrar o valor alvo y_i . Esse valor alvo deve estar contido dentro de uma margem de erro ϵ . Portanto, o hiperplano encontrado é aquele com o coeficiente w , sendo este um vetor normal ao plano, que maximiza o número de pontos dentro da margem de erro, de forma a satisfazer a condição apresentada pela Equação 2.2, enquanto minimiza distância entre o hiperplano e a margem, denotada pela Equação 2.3 (MATHWORKS, 2015; SHARP, 2020).

$$w_i x_i = y_i \quad (2.1)$$

$$|y_i - w_i x_i| \leq \epsilon \quad (2.2)$$

$$\frac{1}{2} \|w\|^2 \quad (2.3)$$

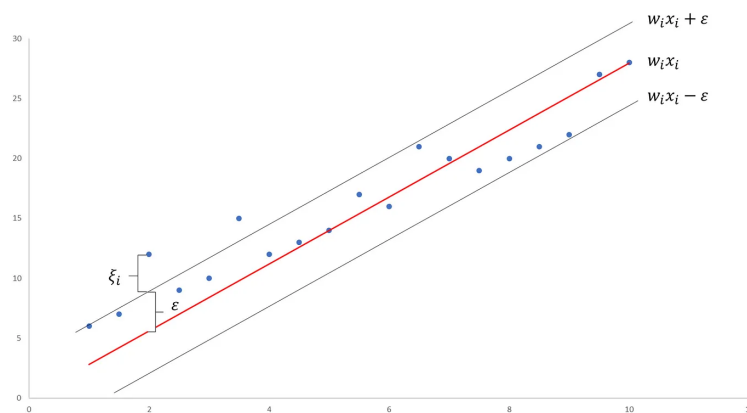
Por melhor que seja o algoritmo, ainda podem haver pontos fora da margem de erro. Todo valor fora desse intervalo ϵ denota um desvio ξ da margem. Levando esse desvio em consideração, as Equações 2.2 e 2.3 se transformam nas Equações 2.4 e 2.5, onde C é uma constante arbitrária proporcional à tolerância do modelo a pontos fora do intervalo de erro (MATHWORKS, 2015; SHARP, 2020).

$$|y_i - w_i x_i| \leq \epsilon + |\xi_i| \quad (2.4)$$

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n |\xi_i| \quad (2.5)$$

A Figura 5 ilustra o algoritmo de SVR.

Figura 5 – Diagrama simplificado do algoritmo de SVR num espaço de 2 dimensões.



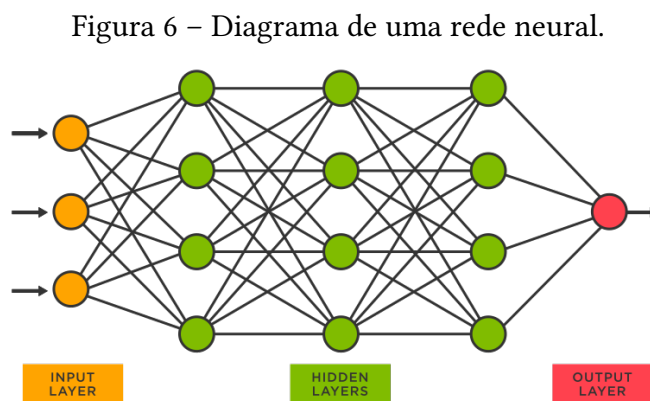
Fonte: (SHARP, 2020).

2.3.3 Redes de Memória Longa de Curto Prazo

Redes de Memória Longa de Curto Prazo, do inglês *Long Short-Term Memory* (LSTM), são uma arquitetura de redes neurais recorrentes.

Uma Rede Neural Artificial (RNA) é uma técnica de AM supervisionado inspirada no cérebro humano. A rede é composta por neurônios, representados por funções matemáticas, que tentam interpretar as entradas recebidas como estímulos sinápticos para produzir uma saída. Esses neurônios estão distribuídos em múltiplas camadas. A camada que recebe diretamente as entradas é chamada de camada de entrada. Essa camada gera estímulos de saída que alimentam as camadas internas da rede, chamadas de camadas ocultas. Esse processo é repetido sucessivas vezes, até que os estímulos chegam na última camada, chamada de camada de saída. Esta é responsável por gerar os resultados finais da rede (RUSSELL; NORVIG, 2009; TIBCO, 2021a).

A Figura 6 traz uma representação visual de uma rede neural.

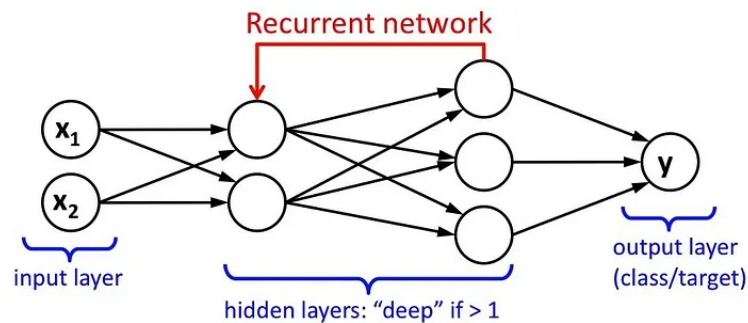


Fonte: (TIBCO, 2021a).

Em uma Rede Neural Recorrente (RNR), o fluxo de informação não segue essa estrutura linear. Os neurônios não recebem informação somente da camada anterior. Elas são capazes de armazenar num estado oculto interno informações geradas em instantes de tempo anteriores para capturar dependências entre diferentes estágios da rede e retroalimentar suas camadas. Por esse motivo, a literatura mostra que as RNRs apresentam um bom potencial para reconhecimento de padrões e previsão de dados sequenciais, como no caso de séries temporais. O problema das RNRs é que essa “memória” armazenada é considerada de curto prazo. Isso ocorre devido ao problema de dissipação do gradiente. À medida que os gradientes calculados pela rede são repassados da camada de saída para a camada de entrada, eles se tornam exponencialmente menores. Dessa forma, os ajustes dos parâmetros da rede se tornam insignificantes e podem fazer com que ela aprenda muito lentamente ou até mesmo nada durante suas iterações seguintes (GOODFELLOW; BENGIO; COURVILLE, 2016; NELSON; PEREIRA; OLIVEIRA, 2017).

A Figura 7 ilustra um exemplo de uma RNR.

Figura 7 – Diagrama de uma RNR.

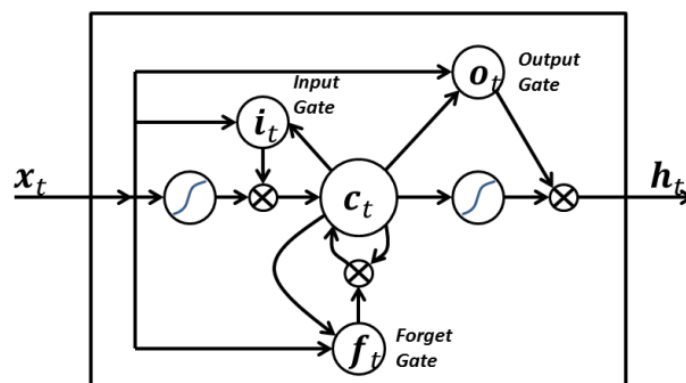


Fonte: (VALKOV, 2017).

As Redes de Memória Longa de Curto Prazo são RNRs que tentam resolver o problema de dissipação do gradiente. Uma rede de LSTM é composta por células de memória que pode armazenar informações sobre os resultados gerados por suas camadas. Elas possuem três mecanismos de porta que regulam o fluxo de entrada e saída de informações desta célula. A porta de esquecimento, que decide quais informações descartar da célula de memória. A porta de entrada, que decide quais informações armazenar na célula de memória. E por fim, a porta de saída, responsável por atualizar o estado oculto da rede com as informações da célula de memória. Assim, o estado oculto da rede é capaz de preservar informações relevantes entre ciclos mais longos, como uma memória de longo prazo (HOCHREITER; SCHMIDHUBER, 1997; NELSON; PEREIRA; OLIVEIRA, 2017; GÉRON, 2017).

A Figura 8 ilustra o diagrama de uma LSTM.

Figura 8 – Diagrama de uma LSTM.



Fonte: (NELSON; PEREIRA; OLIVEIRA, 2017).

2.3.4 Métricas de performance

Neste trabalho, o desempenho das técnicas de AM será avaliado com base em 4 diferentes métricas estatísticas. Essas métricas são comumente utilizadas na literatura para avaliar o desempenho de modelos de AM. São elas: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) e Root Mean Squared Error (RMSE). Termos estes que podem ser traduzidos para português respectivamente como: Erro Absoluto Médio, Erro Absoluto Médio Percentual, Erro Quadrático Médio e Raiz do Erro Quadrático Médio. Para todas as métricas, valores menores indicam que os valores previstos estão mais próximos dos valores esperados. O MAE mede o desvio médio absoluto entre os pontos esperados e previstos. Como os valores absolutos dos produtos analisados variam entre si, é interessante utilizar o MAPE para observar o quanto esse desvio é significativo em termos percentuais. O MSE mede o desvio quadrático médio entre os valores esperados e previstos. Esta métrica é mais sensível a valores muito discrepantes por causa do termo quadrático. A sua raiz pode ser obtida pelo RMSE e comparada ao MAE para que seja observada a diferença dos resultados obtidos ao “penalizar” previsões muito discrepantes. (HYNDMAN; ATHANASOPOULOS, 2021; NASSAR et al., 2020).

O MAE pode ser calculado pela Equação 2.6, o MAPE pode ser calculado pela Equação 2.7, o MSE pode ser calculado pela Equação 2.8 e o RMSE pode ser calculado pela Equação 2.6, Para todas as equações, y_t representa o valor esperado, x_t o valor previsto e N o número de observações.

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - x_t| \quad (2.6)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left| \frac{y_t - x_t}{y_t} \right| \quad (2.7)$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - x_t)^2 \quad (2.8)$$

$$RMSE = \sqrt{MSE} \quad (2.9)$$

2.4 Trabalhos relacionados

Nesta seção estão compilados na Tabela 1 trabalhos relacionados, que também tratam da tarefa de precificação de produtos hortifrutícolas através da análise de séries temporais, e as técnicas de AM utilizadas por cada um.

Reddy (2019), Dharavath e Khosla (2019) e Paredes-Garcia et al. (2019) utilizam os modelos estatísticos ARIMA e SARIMA para realizar previsões de preços de produtos hortifrutícolas em países como Índia e México.

Li, Xu e Li (2010) fazem uma análise comparativa entre o desempenho do modelo estatístico ARIMA com uma RNA na tarefa de predição de preços do tomate na China.

Nassar et al. (2020) fazem uma análise comparativa entre modelos de aprendizado profundo, modelos estatísticos e modelos tradicionais de AM na tarefa de previsão de preços de cinco produtos hortifrutícolas de Taiwan e Canadá. Eles fazem uso de métricas como MAE, MAPE e RMSE para avaliar seus resultados.

Tabela 1 – Trabalhos relacionados.

| Referência | Técnica de AM |
|-------------------------------|--|
| (REDDY, 2019) | SARIMA |
| (DHARAVATH; KHOSLA, 2019) | SARIMA |
| (PAREDES-GARCIA et al., 2019) | ARIMA, SARIMA |
| (NASSAR et al., 2020) | Múltiplas técnicas, incluindo SVR e LSTM |
| (LI; XU; LI, 2010) | RNA, ARIMA |

Fonte: Autoria própria.

3 Metodologia

Este capítulo discorre sobre o processo de execução dos experimentos realizados ao longo do trabalho para o processo de descoberta de conhecimento a partir da base de dados coletada. Neste trabalho, o processo inclui a aquisição dos dados de precificação dos produtos hortifrutícolas, seu pré-processamento e o treinamento das técnicas de AM a partir dos dados gerados.

3.1 Conjunto de dados

O conjunto de dados selecionado para a tarefa de predição foram os preços diários em reais, correspondentes à média simples das cotações diárias realizadas pelos mercados atacadistas, de 10 produtos hortifrutícolas no período de 5 de janeiro de 2015 a 30 de dezembro de 2022 da unidade do CEASA da cidade de São Paulo, o CEAGESP. Os produtos escolhidos foram as principais frutas e hortaliças comercializadas no mês de dezembro de 2022 de acordo com o Boletim Hortigranjeiro (CONAB, 2023) disponibilizado pela Companhia Nacional de Abastecimento (CONAB) em janeiro de 2023, são eles:

- Hortaliças
 - Alface
 - Batata
 - Cebola
 - Cenoura
 - Tomate

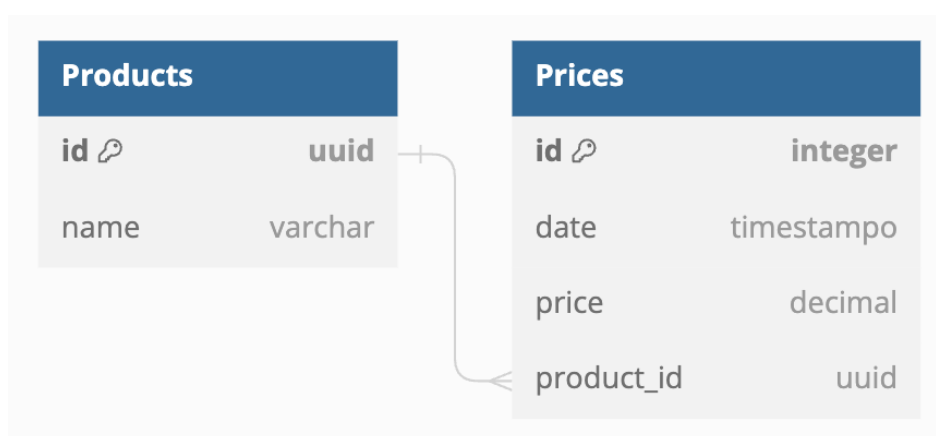
- Frutas
 - Banana
 - Laranja
 - Maçã
 - Mamão
 - Melancia

3.2 Aquisição e preparação dos dados

Os dados foram obtidos do sistema do Programa Brasileiro de Modernização do Mercado Hortigranjeiro (Prohort)¹.

Foi realizada uma consulta no sistema para cada produto selecionado. O resultado de cada consulta foi o conjunto de preços diários de cada produto no período de tempo definido. Esses resultados foram exportados em arquivos no formato CSV e, com o auxílio de *scripts* nas linguagens de programação Python² e JavaScript³, processados e armazenados em um banco de dados MySQL⁴ na nuvem. A Figura 9 apresenta a estrutura das duas tabelas criadas no banco, uma tabela de produtos com uma relação “um para muitos” com uma tabela de preços.

Figura 9 – Esquema das tabelas do banco de dados.



Fonte: Autoria própria.

Para fins de demonstração, as etapas seguintes discorrem sobre a análise e predição de preços do produto Tomate, considerando que o mesmo processo pode ser generalizados para obter os mesmos resultados dos demais produtos. Os resultados de todas as análises podem ser conferidos no Apêndice A.

O primeiro passo foi a preparação dos dados para a mineração pelos modelos de AM. Foi observado que o conjunto de dados possui apenas 1626 observações, enquanto o número de observações diárias esperado para o intervalo de tempo escolhido era de 2917, pois é o número de dias entre as datas escolhidas. Isso é decorrente do fato de que as Centrais de Abastecimento só realizam a coleta de preços nos dias considerados de “forte comercialização”, de acordo com as informações do Sistema Prohort. Não podem haver instantes de tempo sem observações para se realizar a predição dos dados, portanto, utilizando interpolação

¹ <http://dw.ceasa.gov.br/>

² <https://www.python.org/>

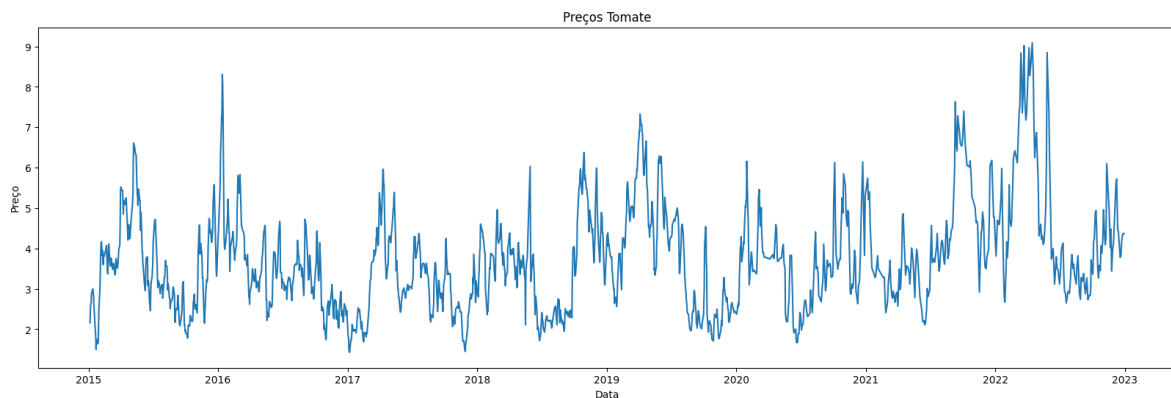
³ <https://developer.mozilla.org/en-US/docs/Web/JavaScript>

⁴ <https://www.mysql.com/>

linear, foi possível traçar uma reta entre os dois pontos de observação mais próximos de cada instante sem registro para estimar o valor das observações que não foram registradas.

O próximo passo foi fazer uma análise gráfica da série resultante. A Figura 10 apresenta o gráfico da série temporal dos preços diários do Tomate com os valores faltantes já aproximados por interpolação linear.

Figura 10 – Série temporal dos preços diários do Tomate.



Fonte: Autoria própria.

Para alguns produtos, foi possível observar graficamente preços registrados com o valor de R\$0,00 e preços com um valor 10 vezes maior do que a média dos dados. Essas observações foram consideradas como *outliers*⁵, removidas e aproximadas por meio de interpolação linear, considerando o mesmo princípio detalhado na etapa anterior.

3.3 Mineração dos dados pelos modelos de AM

O primeiro passo foi estruturar a base de dados para um conjunto de aprendizado supervisionado. Foi definido um horizonte de tempo de sete dias, para serem utilizados como atributo preditivo pelos algoritmos estudados. Estes são os dados disponibilizados para a indução do modelo pelas técnicas de AM, já o dia seguinte a estes sete é o atributo alvo, ou seja, o valor desejado para predição dos modelos. A Figura 11 apresenta o esquema utilizado para a estruturação dos dados como um conjunto de aprendizado supervisionado.

⁵ Observações atípicas, discrepantes

Figura 11 – Esquema do conjunto de aprendizado supervisionado.

| Dia | Preço do produto (R\$) | Preço alvo (R\$) |
|-----|------------------------|------------------|
| 1 | 2,00 | - |
| 2 | 2,07 | - |
| 3 | 2,08 | - |
| 4 | 2,05 | - |
| 5 | 2,03 | - |
| 6 | 1,98 | - |
| 7 | 2,03 | - |
| 8 | 2,06 | 2,06 |
| 9 | 2,07 | 2,07 |
| 10 | 2,09 | 2,09 |

Fonte: Autoria própria.

Neste trabalho, foram geradas previsões para 4 períodos de tempo futuro diferentes: 7, 30, 90 e 180 dias. Os modelos foram treinados com um conjunto de dados de treinamento correspondentes a ao número total de observações decrescido em 180, correspondente ao maior período para o qual serão geradas as previsões.

3.3.1 Parâmetros de configuração das técnicas de AM

São apresentados a seguir, os parâmetros utilizados para a configuração das técnicas de AM aplicadas na tarefa de previsão. Para a implementação das técnicas de RF e SVR foram utilizados os métodos configuráveis disponíveis na biblioteca scikit-learn⁶. Para LSTM, foi utilizado o pacote Keras⁷, uma API de alto nível para construir e treinar modelos usando pacote TensorFlow⁸.

- RF – Número de árvores de decisão da floresta, $n_estimators=100$, a função para medir a qualidade de uma divisão foi o erro quadrático médio, $criterion="squared_error"$, profundidade máxima de uma árvore de decisão individual, $max_depth=10$, número mínimo de amostras necessárias para dividir um nó interno, $min_samples_split=5$, número mínimo de amostras necessárias para estar em um nó folha, $min_samples_leaf=5$.
- SVR - Parâmetro de regularização, $C=0.001$, função kernel que mapeia os dados de entrada para um espaço de mais dimensões, $kernel='rbf'$, coeficiente da função kernel, $gamma=0.1$, valor ϵ da margem de erro, $epsilon=0.1$.

⁶ <https://scikit-learn.org/stable/>

⁷ <https://keras.io/>

⁸ <https://www.tensorflow.org/>

- LSTM - Foi utilizado um modelo sequencial para gerar uma pilha linear de camadas. Primeiramente uma camada LSTM com 50 células, *units=50*, seguida de uma rede neural densa que será a camada de saída com uma célula, *units=1*. Para a compilação do modelo foi definido o tipo de otimizador, *optimizer="adam"*, a função de perda como sendo o erro quadrático médio, *loss="mean_squared_error"*, e a métrica de validação do treinamento como sendo o erro quadrático médio, *metrics=[tf.keras.metrics.MeanSquaredError()]*.

Com a configuração apresentada, os modelos foram treinados e tarefa de previsão para os 4 períodos propostos foi realizada.

4 Análise experimental e resultados

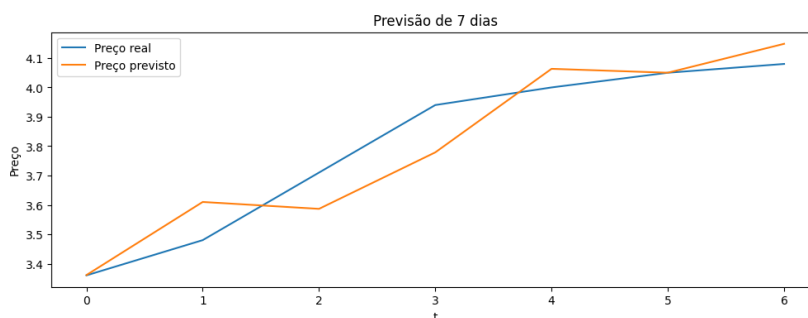
Neste capítulo é feita a avaliação e comparação dos resultados preditivos obtidos a partir de cada um dos modelos de AM, utilizando métricas estatísticas de performance e simulando um cenário hipotético de vendas considerando o conhecimento gerado pelos modelos.

4.1 Previsões geradas

São apresentados a seguir, os gráficos das séries temporais previstas para o preço do tomate utilizando cada um dos algoritmos de AM selecionados.

As Figuras 12, 13, 14, 15 apresentam as séries temporais de valores reais e as séries temporais de valores previstos pelo algoritmo de RF para os períodos de 7, 30, 90 e 180 dias, respectivamente. É possível observar pelos gráficos uma diferença no resultado obtido pelas previsões, apesar de todas terem sido feitas pelo mesmo modelo treinado. As séries previstas nos gráficos das Figuras 12 e 13 aparentam ter se desviado mais dos valores reais, enquanto as séries previstas nas Figuras 14 e 15 quase sobrepõem a série de valores esperados. Contudo, é importante ressaltar que esse efeito pode ter sido provocado pelo aumento da escala dos gráficos à medida que o período da previsão aumenta, o que torna esses desvios invisíveis caso sejam suficientemente pequenos.

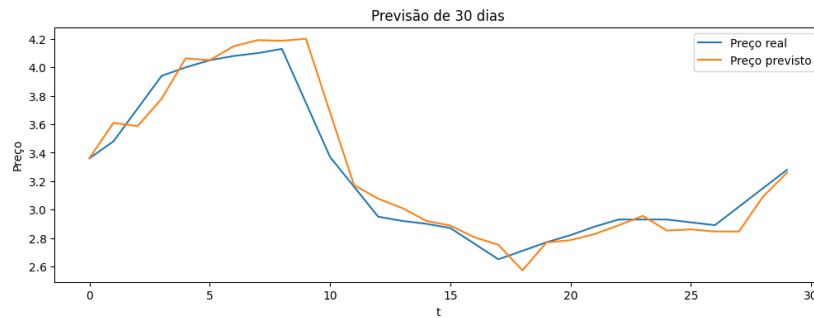
Figura 12 – Previsão de 7 dias utilizando RF.



Fonte: Autoria própria.

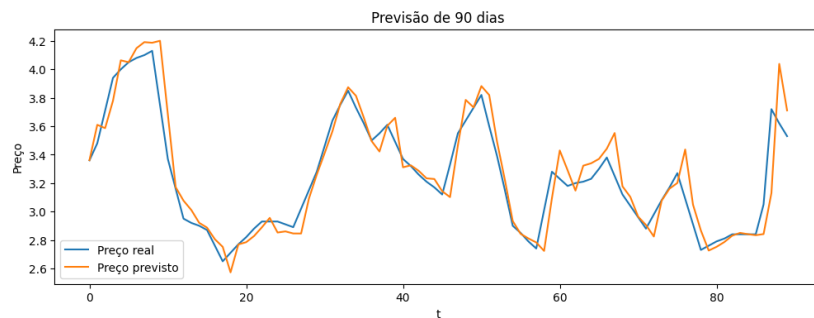
As Figuras 16, 17, 18, 19 apresentam as séries temporais de valores reais e as séries temporais de valores previstos pelo algoritmo de SVR para os períodos de 7, 30, 90 e 180 dias, respectivamente. De maneira semelhante ao algoritmo de RF, é possível observar pelos gráficos uma diferença no resultado obtido pelas previsões, apesar de todas terem sido feitas pelo mesmo modelo treinado. As séries previstas para os períodos mais curtos possuem desvios visíveis em comparação às séries de valores reais. Os desvios vão diminuindo à

Figura 13 – Previsão de 30 dias utilizando RF.



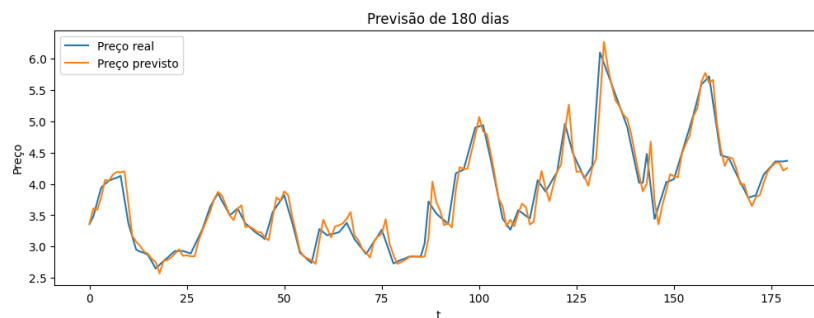
Fonte: Autoria própria.

Figura 14 – Previsão de 90 dias utilizando RF.



Fonte: Autoria própria.

Figura 15 – Previsão de 180 dias utilizando RF.

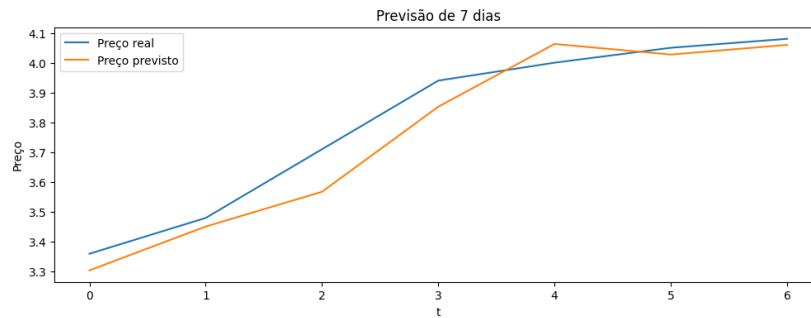


Fonte: Autoria própria.

medida que o período da previsão aumenta. Porém, como observado anteriormente, este efeito pode ser decorrente do aumento da escala dos gráficos e é preciso fazer mais análises para se chegar a alguma conclusão sobre o desempenho dos modelos.

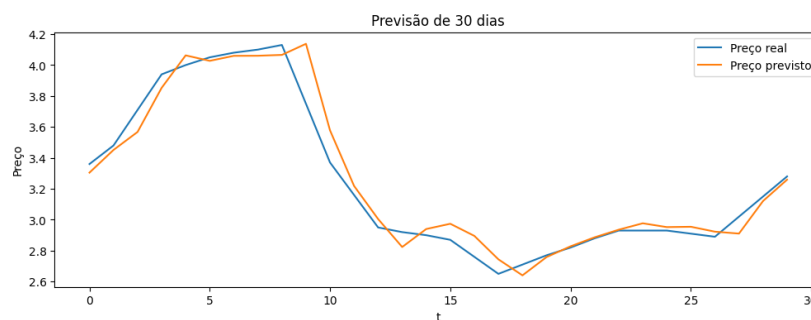
As Figuras 20, 21, 22, 23 apresentam as séries temporais de valores reais e as séries temporais de valores previstos pelo algoritmo de LSTM para os períodos de 7, 30, 90 e 180 dias, respectivamente. Mais uma vez, observa-se um comportamento semelhante aos gráficos dos algoritmos de RF e SVR. As séries previstas para os períodos mais curtos possuem desvios

Figura 16 – Previsão de 7 dias utilizando SVR.



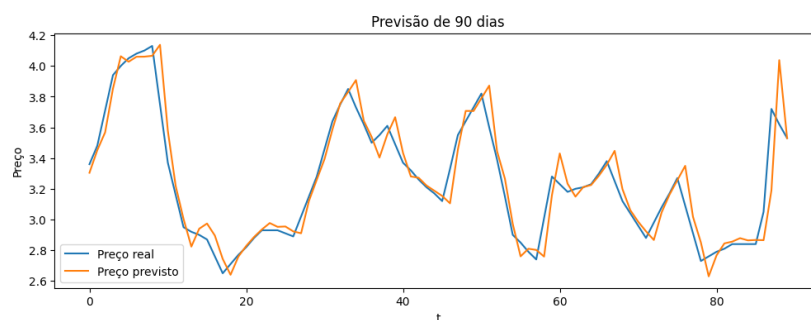
Fonte: Autoria própria.

Figura 17 – Previsão de 30 dias utilizando SVR.



Fonte: Autoria própria.

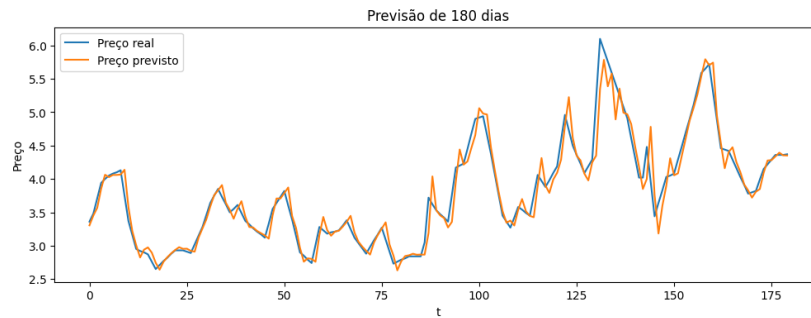
Figura 18 – Previsão de 90 dias utilizando SVR.



Fonte: Autoria própria.

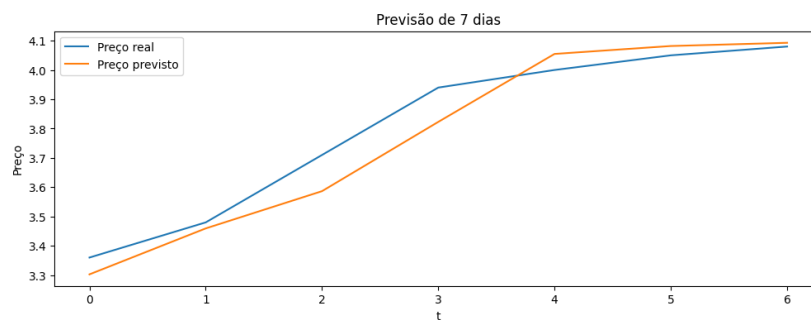
visíveis em comparação às séries de valores reais. À medida que o período da previsão aumenta, as séries de valores reais e previstos quase se sobrepõem. Por causa da mudança da escala dos gráficos, é necessário fazer mais análises para se chegar a alguma conclusão em relação ao desempenho dos modelos para diferentes intervalos de tempo.

Figura 19 – Previsão de 180 dias utilizando SVR.



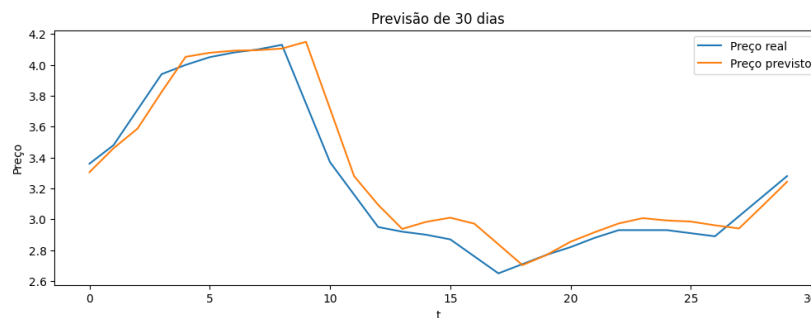
Fonte: Autoria própria.

Figura 20 – Previsão de 7 dias utilizando LSTM.



Fonte: Autoria própria.

Figura 21 – Previsão de 30 dias utilizando LSTM.

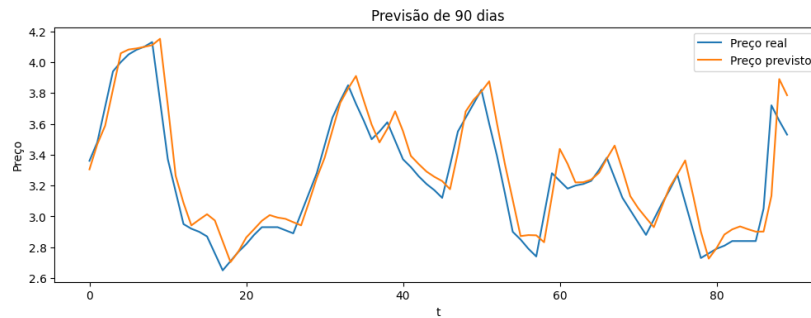


Fonte: Autoria própria.

4.2 Avaliação de desempenho preditivo dos modelos de AM

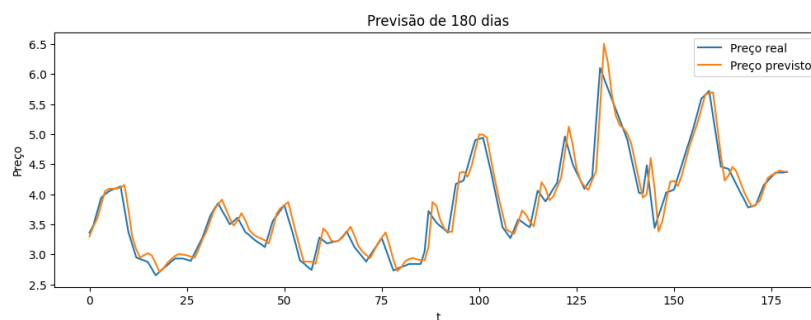
Com o resultado obtido pelas previsões de cada um dos modelos para os períodos de tempo propostos, o desempenho dos modelos foi avaliado por meio da comparação de métricas estatísticas e de uma simulação de um cenário hipotético, em que um produtor utiliza os modelos obtidos para aumentar seus lucros.

Figura 22 – Previsão de 90 dias utilizando LSTM.



Fonte: Autoria própria.

Figura 23 – Previsão de 180 dias utilizando LSTM.



Fonte: Autoria própria.

4.2.1 Avaliação a partir de métricas estatísticas

A partir das observações preditivas obtidas na etapa anterior, foram calculadas o valor das quatro métricas estatísticas de desempenho selecionadas para cada um dos períodos analisados e cada um dos métodos estudados.

A Tabela 2 apresenta os valores das 4 métricas de performance calculadas para as previsões do período de 7 dias de cada um dos modelos estudados. De acordo com os resultados obtidos, todas as métricas apontam que o modelo com melhor desempenho foi o SVR.

A Tabela 3 apresenta os valores das 4 métricas de performance calculadas para as previsões do período de 30 dias de cada um dos modelos estudados. De acordo com os resultados obtidos, todas as métricas apontam que o modelo com melhor desempenho foi o SVR.

A Tabela 4 apresenta os valores das 4 métricas de performance calculadas para as previsões do período de 90 dias de cada um dos modelos estudados. De acordo com os resultados obtidos, todas as métricas apontam que o modelo com melhor desempenho foi o SVR.

A Tabela 5 apresenta os valores das 4 métricas de performance calculadas para as previsões do período de 180 dias de cada um dos modelos estudados. De acordo com os resultados obtidos, todas as métricas apontam que o modelo com melhor desempenho foi o RF.

Tabela 2 – Métricas de performance para a previsão de 7 dias do Tomate.

| Métrica | Modelo | Valor |
|---------|--------|--------|
| MAE | RF | 0.078 |
| MAE | SVR | 0.06 |
| MAE | LSTM | 0.059 |
| MAPE | RF | 0.021% |
| MAPE | SVR | 0.016% |
| MAPE | LSTM | 0.016% |
| MSE | RF | 0.01 |
| MSE | SVR | 0.005 |
| MSE | LSTM | 0.005 |
| RMSE | RF | 0.098 |
| RMSE | SVR | 0.072 |
| RMSE | LSTM | 0.072 |

Fonte: Autoria própria.

Tabela 3 – Métricas de performance para a previsão de 30 dias do Tomate.

| Métrica | Modelo | Valor |
|---------|--------|--------|
| MAE | RF | 0.086 |
| MAE | SVR | 0.07 |
| MAE | LSTM | 0.092 |
| MAPE | RF | 0.026% |
| MAPE | SVR | 0.022% |
| MAPE | LSTM | 0.029% |
| MSE | RF | 0.016 |
| MSE | SVR | 0.011 |
| MSE | LSTM | 0.017 |
| RMSE | RF | 0.127 |
| RMSE | SVR | 0.102 |
| RMSE | LSTM | 0.132 |

Fonte: Autoria própria.

Tabela 4 – Métricas de performance para a previsão de 90 dias do Tomate.

| Métrica | Modelo | Valor |
|---------|--------|--------|
| MAE | RF | 0.094 |
| MAE | SVR | 0.082 |
| MAE | LSTM | 0.112 |
| MAPE | RF | 0.028% |
| MAPE | SVR | 0.025% |
| MAPE | LSTM | 0.035% |
| MSE | RF | 0.02 |
| MSE | SVR | 0.015 |
| MSE | LSTM | 0.023 |
| RMSE | RF | 0.142 |
| RMSE | SVR | 0.124 |
| RMSE | LSTM | 0.151 |

Fonte: Autoria própria.

Tabela 5 – Métricas de performance para a previsão de 180 dias do Tomate.

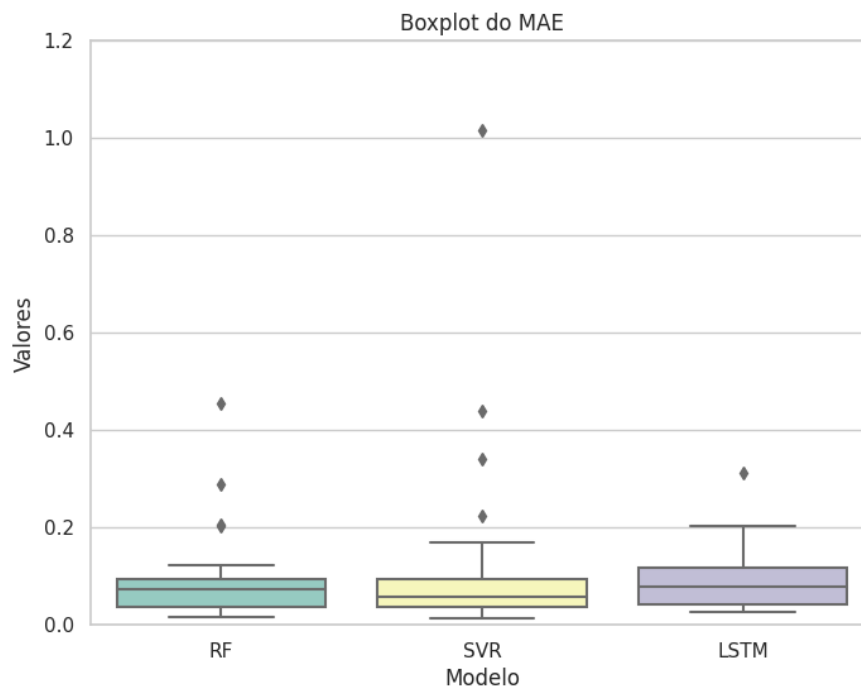
| Métrica | Modelo | Valor |
|---------|--------|--------|
| MAE | RF | 0.121 |
| MAE | SVR | 0.121 |
| MAE | LSTM | 0.146 |
| MAPE | RF | 0.031% |
| MAPE | SVR | 0.031% |
| MAPE | LSTM | 0.038% |
| MSE | RF | 0.033 |
| MSE | SVR | 0.036 |
| MSE | LSTM | 0.04 |
| RMSE | RF | 0.183 |
| RMSE | SVR | 0.189 |
| RMSE | LSTM | 0.2 |

Fonte: Autoria própria.

Para observar o intervalo de distribuição do valor das métricas de cada modelo, foram gerados gráficos do tipo box-plot que agregam os resultados de cada uma das métricas para todos os produtos e intervalos de tempo analisados separados por modelo.

A Figura 24 apresenta o gráfico box-plot do MAE para cada um dos modelos. É possível observar que o modelo de SVR foi o que apresentou um maior número de *outliers*. Apesar disso, a mediana, os quartis e os limites dos gráficos não apresentam uma variação significativa.

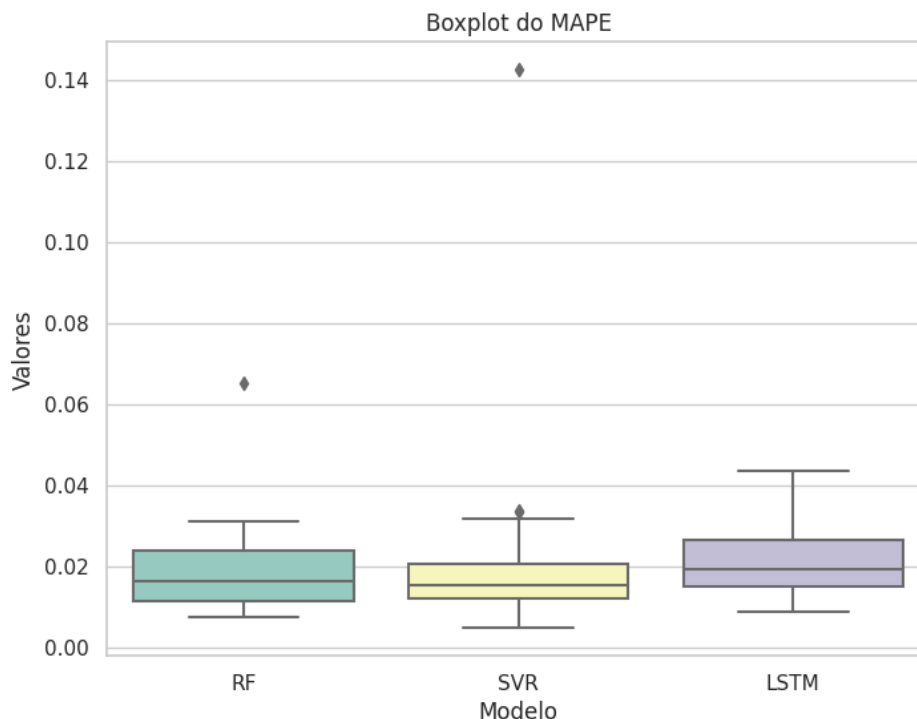
Figura 24 – Gráfico de box-plot para os valores do MAE de cada algoritmo de AM estudado.



Fonte: Autoria própria.

A Figura 25 apresenta o gráfico box-plot do MAPE para cada um dos modelos. É possível observar que o modelo de SVR foi o que apresentou um maior número de *outliers*. O modelo de SVR também apresentou uma performance ligeiramente mais consistente que os demais, pois possui a menor distância entre o primeiro e o terceiro quartil indicada pelo gráfico. Porém, a distribuição dos valores, mais uma vez, é muito semelhante.

Figura 25 – Gráfico de box-plot para os valores do MAPE de cada algoritmo de AM estudado.

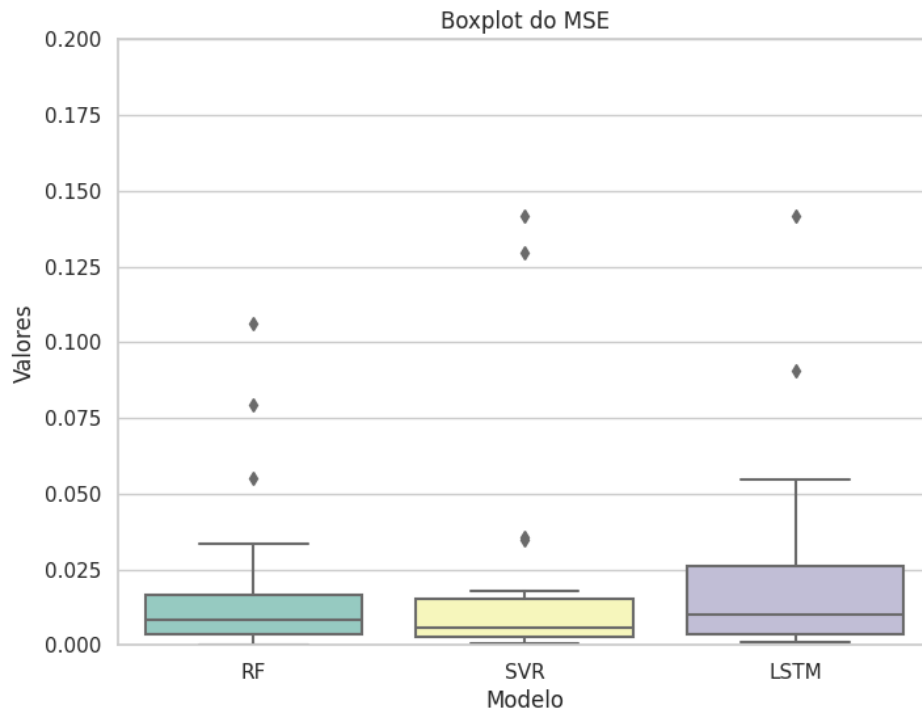


Fonte: Autoria própria.

A Figura 26 apresenta o gráfico box-plot do MSE para cada um dos modelos. A escala do gráfico precisou ser alterada para permitir uma melhor visualização, o que fez com que alguns *outliers* não coubessem na imagem. Contudo, os algoritmos de RF e SVR apresentaram o maior número de *outliers*, ambos com cinco. O modelo de LSTM apresentou uma performance ligeiramente inferior aos demais, pois possui uma distância maior entre o primeiro e o terceiro quartil indicada pelo gráfico. Porém, novamente, a diferença não chega a ser muito significativa.

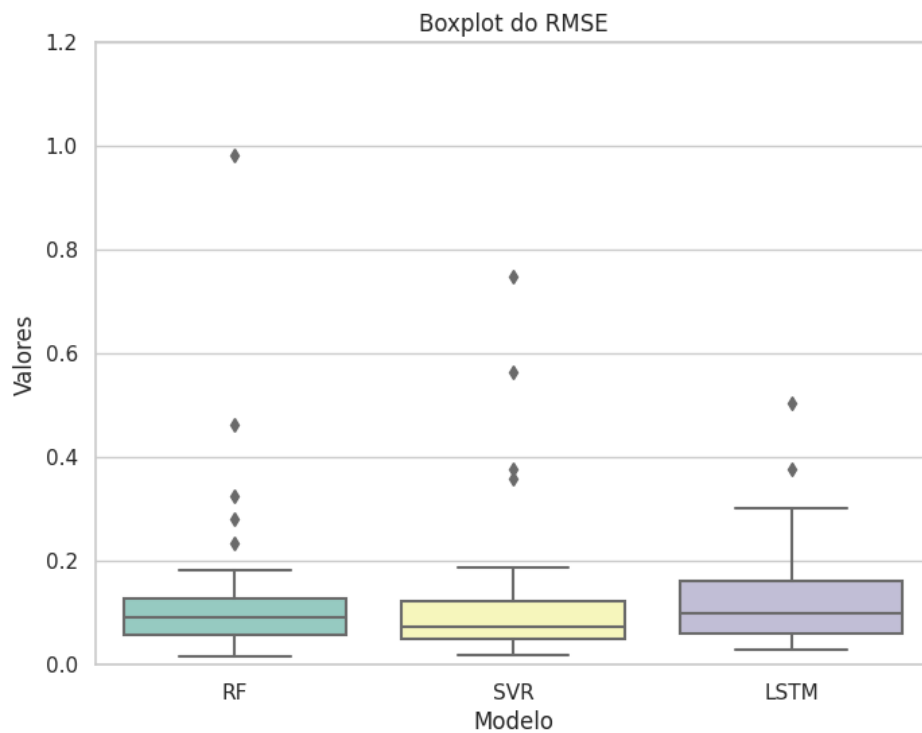
A Figura 27 apresenta o gráfico box-plot do RMSE para cada um dos modelos. A escala do gráfico precisou ser alterada para permitir uma melhor visualização, o que fez com que alguns *outliers* não coubessem na imagem. Contudo, os algoritmos de RF e SVR apresentaram o maior número de *outliers*, ambos com cinco. Uma última vez, constatou-se que a mediana, os quartis e os limites dos gráficos não apresentam uma variação significativa.

Figura 26 – Gráfico de box-plot para os valores do MSE de cada algoritmo de AM estudado.



Fonte: Autoria própria.

Figura 27 – Gráfico de box-plot para os valores do RMSE de cada algoritmo de AM estudado.



Fonte: Autoria própria.

4.2.2 Simulação de desempenho financeiro

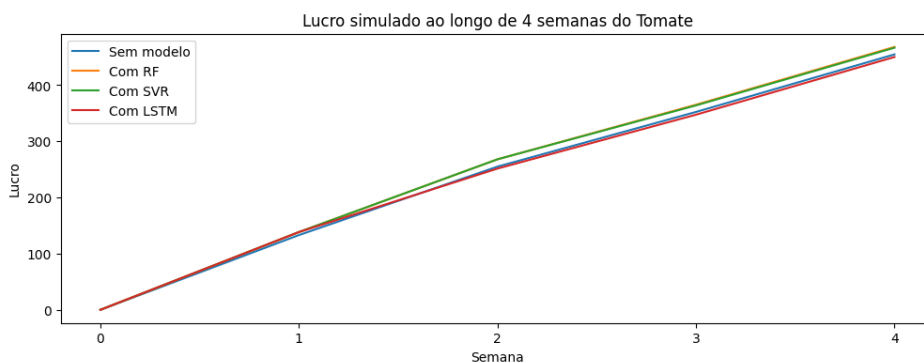
A seguir, foi feita uma simulação de um cenário hipotético, em que um produtor utiliza o conhecimento obtido pelas previsões dos modelos treinados, para maximizar seu lucro na venda do produto X no período de 28 dias, ou quatro semanas.

Supondo que o produtor possui um estoque de 35kg para ser negociado durante cada intervalo de sete dias, será simulada a diferença no montante final obtido pelas vendas da mercadoria em um cenário onde ela é negociada em um volume uniformemente distribuído ao longo dos dias, na situação em que o produtor desconhece quais são os picos de preço na semana, e em um cenário onde ele negocia o mesmo volume total de produtos distribuído uniformemente ao longo de cinco dias utilizando o conhecimento de cada um dos modelos.

Estes cinco dias serão os instantes de tempo com os maiores valores de preço observados em previsões de cada um dos modelos. Nesta tarefa, cada modelo prevê o preço do dia seguinte com base nos dados dos últimos sete dias e utiliza seu próprio resultado para gerar previsões para os próximos sete dias. O processo é repetido ao final de cada semana.

A Figura 28 apresenta o gráfico da diferença do lucro acumulado em reais ao longo de 4 semanas do produtor que vendeu o produto Tomate sem o auxílio do modelo e com o auxílio dos modelos.

Figura 28 – Gráfico dos resultados da simulação com o produto Tomate.



Fonte: Autoria própria.

A simulação mostrou que o lucro do produtor com o conhecimento gerado pelo modelo de RF foi 2,93% maior, pelo modelo de SVR foi 2,62% maior e pelo modelo de LSTM foi 1,01% menor. As Tabelas 6, 7, 8 apresentam o aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando os modelos de RF, SVR e LSTM, respectivamente. Em média, o modelo de RF gerou um aumento nos lucros de 0,29%, o modelo de SVR gerou um aumento de 0,74% e o modelo de LSTM gerou um aumento de 0,34%.

Tabela 6 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando RF.

| Produto | Aumento no lucro |
|----------|------------------|
| Alface | -0,11% |
| Banana | 1,21% |
| Batata | 0,60% |
| Cebola | 0,50% |
| Cenoura | 0,37% |
| Laranja | -1,10% |
| Maçã | -0,03% |
| Mamão | -1,43% |
| Melancia | -0,08% |
| Tomate | 2,93% |

Fonte: A autoria própria.

Tabela 7 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando SVR.

| Produto | Aumento no lucro |
|----------|------------------|
| Alface | -0,63% |
| Banana | 0,61% |
| Batata | 0,46% |
| Cebola | 1,04% |
| Cenoura | 0,46% |
| Laranja | 0,13% |
| Maçã | -0,17% |
| Mamão | 2,43% |
| Melancia | 0,46% |
| Tomate | 2,62% |

Fonte: A autoria própria.

Tabela 8 – Aumento percentual do lucro do produtor para cada um dos produtos analisados utilizando LSTM.

| Produto | Aumento no lucro |
|----------|------------------|
| Alface | -0,11% |
| Banana | 1,03% |
| Batata | 0,02% |
| Cebola | 1,25% |
| Cenoura | 0,58% |
| Laranja | -0,52% |
| Maçã | -0,09% |
| Mamão | 2,43% |
| Melancia | -0,15% |
| Tomate | -1,01% |

Fonte: A autoria própria.

5 Conclusão

As análises realizadas sobre os desempenhos obtidos pelas técnicas de AM treinadas a partir de séries temporais contendo o preço de produtos hortifrutícolas comercializados entre janeiro de 2015 e dezembro de 2022 na unidade do CEAGESP da cidade de São Paulo, geraram indícios de que as técnicas de AM analisadas apresentam um desempenho muito semelhante na tarefa avaliada. Em média, o algoritmo de SVR proporcionou resultados preditivos marginalmente melhores quando comparada às técnicas de RF e LSTM, de acordo com as previsões realizadas neste trabalho. Por fim, o resultado obtido através da simulação de desempenho financeiro mostra que, apesar da média dos resultados ser positiva, os resultados obtidos pelos modelos não foram consistentes, por isso não é possível concluir se os modelos têm potencial para servirem como ferramentas que auxiliam o pequeno produtor a tomar melhores decisões sobre a gestão de vendas de sua mercadoria.

É importante ressaltar que, com mais dados, poderiam ser feitas análises de performance para mais produtos e simulações para outros períodos de tempo, permitindo assim investigar se a configuração selecionada dos modelos realmente é a que melhor generaliza o conhecimento obtido pelos dados de treinamento para a tarefa de predição.

Assim sendo, algumas proposições válidas para trabalhos futuros são utilizar Aprendizado de Máquina Automático para automatizar a seleção de parâmetros dos modelos, fazer análises utilizando mais produtos e mais modelos, como ARIMA ou SARIMA, e fazer previsões para horizontes de tempo maiores utilizando os resultados dos próprios modelos, para auxiliar produtores em momentos além da negociação de mercadorias.

Referências

- ABRACEN, A. B. de Centrais de A. roposta da Abracen para plano do governo federal na área de abastecimento alimentar. 2007. Disponível em: <https://www.conab.gov.br/info-agro/hortigranjeiros-prohort/publicacoes-do-setor-hortigranjeiro/item/download/592_3d0a8e195763f45df452c716b2cc9663>. Acesso em: 20 jul. 2023.
- BARRETO JÚNIOR, D. Análise crítica do desempenho das Centrais de Abastecimento – CEASAS. *Anais da Academia Pernambucana de Ciência Agronômica*, v. 1, p. 20–26, ago. 2013. Disponível em: <<https://www.journals.ufrpe.br/index.php/apca/article/view/40>>. Acesso em: 20 jul. 2023.
- BRASIL. Lei nº 11.326, de 24 de julho de 2006. estabelece as diretrizes para a formulação da Política Nacional da Agricultura Familiar e Empreendimentos Familiares Rurais. *Diário Oficial da União*, Brasília, DF, 2006. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/11326.htm>. Acesso em: 20 jul. 2023.
- BREIMAN, L. Random forests. *Machine Learning*, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. [S.l.]: Springer International Publishing, 2016. (Springer Texts in Statistics).
- CEASA/CE, C. de Abastecimento do Ceará S. *Perguntas Frequentes*. 2016. Disponível em: <<https://www.ceasa-ce.com.br/aceso-a-informacao/perguntas-frequentes/>>. Acesso em: 20 jul. 2023.
- CONAB, C. N. de A. *Boletim Hortigranjeiro Janeiro 2023*. 2023. Disponível em: <https://www.conab.gov.br/info-agro/hortigranjeiros-prohort/boletim-hortigranjeiro/item/download/46155_75fdc0987ae8bda1c19a7895add1ac79>.
- DHARAVATH, R.; KHOSLA, E. Seasonal arima to forecast fruits and vegetable agricultural prices. In: *2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)*. [S.l.: s.n.], 2019. p. 47–52.
- DIEESE, D. Intersindical de Estatística e Estudos Socioeconômicos. *Estimativas: Agricultura Familiar e Renda Básica*. 2020. Disponível em: <<https://www.dieese.org.br/outraspublicacoes/2020/agriculturaFamiliarBrasilEstados.html>>. Acesso em: 20 jul. 2023.
- DRUCKER, H. et al. Support vector regression machines. MIT Press, v. 9, 1996.
- Embrapa, E. B. de P. A. *Perguntas e respostas | Agricultura Familiar*. 2021. Disponível em: <<https://www.embrapa.br/tema-agricultura-familiar/perguntas-e-respostas>>. Acesso em: 20 jul. 2023.
- FLACH, P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. [S.l.]: Cambridge University Press, 2012.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.

GÉRON, A. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1491962299.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 0899-7667.

HYNDMAN, R.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 3rd. ed. Australia: OTexts, 2021.

IBGE, I. B. de Geografia e E. *Censo Agro 2017: população ocupada nos estabelecimentos agropecuários cai 8,8%*. 2019. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/25789-censo-agro-2017-populacao-ocupada-nos-estabelecimentos-agropecuarios-cai-8-8>>. Acesso em: 20 jul. 2023.

JAVATPOINT. *Support Vector Machine Algorithm*. 2019. Disponível em: <<https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>>. Acesso em: 30 jul. 2023.

KHAIDEM, L.; SAHA, S.; DEY, S. R. Predicting the direction of stock market prices using random forest. *CoRR*, abs/1605.00003, 2016.

LI, G. qiong; XU, S. wei; LI, Z. min. Short-term price forecasting for agro-products using artificial neural networks. *Agriculture and Agricultural Science Procedia*, v. 1, p. 278–287, 2010. International Conference on Agricultural Risk and Food Security 2010.

MATHWORKS. *Understanding Support Vector Machine Regression*. 2015. Disponível em: <<https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html>>. Acesso em: 29 jul. 2023.

MATHWORKS. *Support Vector Machines for Binary Classification*. 2016. Disponível em: <<https://de.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html>>. Acesso em: 29 jul. 2023.

MCCARTHY, J. What is artificial intelligence? 2004. Disponível em: <<https://www-formal.stanford.edu/jmc/whatisai.pdf>>. Acesso em: 21 jul. 2023.

NASSAR, L. et al. Deep learning based approach for fresh produce market price prediction. p. 1–7, 2020.

NELSON, D. M. Q.; PEREIRA, A. C. M.; OLIVEIRA, R. A. de. Stock market's price movement prediction with lstm neural networks. p. 1419–1426, 2017.

PAREDES-GARCIA, W. J. et al. Price forecasting and span commercialization opportunities for mexican agricultural products. *Agronomy*, v. 9, n. 12, 2019. ISSN 2073-4395.

REDDY, A. A. Price forecasting of tomatoes. *International Journal of Vegetable Science*, Taylor Francis, v. 25, n. 2, p. 176–184, 2019.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: a modern approach*. 3. ed. [S.l.]: Pearson, 2009.

SHARP, T. *An Introduction to Support Vector Regression (SVR)*. 2020. Disponível em: <<https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>>. Acesso em: 30 jul. 2023.

TAY, F.; CAO, L. Application of support vector machines in financial time series forecasting. *Omega*, v. 29, p. 309–317, 02 2001.

TIBCO. *What is a Neural Network?* 2021. Disponível em: <<https://www.tibco.com/reference-center/what-is-a-neural-network>>. Acesso em: 30 jul. 2023.

TIBCO. *What is a Random Forest?* 2021. Disponível em: <<https://www.tibco.com/reference-center/what-is-a-random-forest>>. Acesso em: 23 jul. 2023.

TURING, A. M. Computing machinery and intelligence. *Mind*, Oxford University Press on behalf of the Mind Association, v. 59, n. 236, p. 433–460, 1950.

UBEDA, G. das N. Estudo comparativo entre técnicas de aprendizado de máquina tradicionais e de aprendizado profundo para a previsão do preço de ações listadas na bolsa de valores brasileira. Instituto de Ciências e Engenharia, Universidade Estadual Paulista "Júlio de Mesquita Filho", 2022.

VALKOV, V. *Making a Predictive Keyboard using Recurrent Neural Networks*. 2017. Disponível em: <<https://venelinvalkov.medium.com/making-a-predictive-keyboard-using-recurrent-neural-networks-tensorflow-for-hackers-part-v-3f238d824218>>. Acesso em: 30 jul. 2023.

A Apêndices

Está contido neste apêndice, as tabelas com todos os resultados obtidos ao longo dos experimentos para o referido Trabalho de Conclusão de Curso.

Tabela 9 – Métricas de performance das previsões do produto Alfaca

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.094 |
| 7 dias | MAE | SVR | 0.168 |
| 7 dias | MAE | LSTM | 0.076 |
| 7 dias | MAPE | RF | 0.008% |
| 7 dias | MAPE | SVR | 0.014% |
| 7 dias | MAPE | LSTM | 0.006% |
| 7 dias | MSE | RF | 0.014 |
| 7 dias | MSE | SVR | 0.035 |
| 7 dias | MSE | LSTM | 0.008 |
| 7 dias | RMSE | RF | 0.118 |
| 7 dias | RMSE | SVR | 0.187 |
| 7 dias | RMSE | LSTM | 0.089 |
| 30 dias | MAE | RF | 0.203 |
| 30 dias | MAE | SVR | 0.34 |
| 30 dias | MAE | LSTM | 0.198 |
| 30 dias | MAPE | RF | 0.016% |
| 30 dias | MAPE | SVR | 0.026% |
| 30 dias | MAPE | LSTM | 0.015% |
| 30 dias | MSE | RF | 0.106 |
| 30 dias | MSE | SVR | 0.319 |
| 30 dias | MSE | LSTM | 0.125 |
| 30 dias | RMSE | RF | 0.326 |
| 30 dias | RMSE | SVR | 0.564 |
| 30 dias | RMSE | LSTM | 0.353 |
| 90 dias | MAE | RF | 0.205 |
| 90 dias | MAE | SVR | 0.224 |
| 90 dias | MAE | LSTM | 0.178 |
| 90 dias | MAPE | RF | 0.018% |
| 90 dias | MAPE | SVR | 0.019% |
| 90 dias | MAPE | LSTM | 0.016% |
| 90 dias | MSE | RF | 0.079 |
| 90 dias | MSE | SVR | 0.142 |
| 90 dias | MSE | LSTM | 0.083 |
| 90 dias | RMSE | RF | 0.281 |
| 90 dias | RMSE | SVR | 0.376 |
| 90 dias | RMSE | LSTM | 0.289 |
| 180 dias | MAE | RF | 0.289 |
| 180 dias | MAE | SVR | 0.439 |
| 180 dias | MAE | LSTM | 0.321 |
| 180 dias | MAPE | RF | 0.024% |
| 180 dias | MAPE | SVR | 0.034% |
| 180 dias | MAPE | LSTM | 0.026% |
| 180 dias | MSE | RF | 0.215 |
| 180 dias | MSE | SVR | 0.559 |
| 180 dias | MSE | LSTM | 0.268 |
| 180 dias | RMSE | RF | 0.464 |
| 180 dias | RMSE | SVR | 0.748 |
| 180 dias | RMSE | LSTM | 0.518 |

Fonte: Autoria própria.

Tabela 10 – Métricas de performance das previsões do produto Banana

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.025 |
| 7 dias | MAE | SVR | 0.024 |
| 7 dias | MAE | LSTM | 0.03 |
| 7 dias | MAPE | RF | 0.008% |
| 7 dias | MAPE | SVR | 0.008% |
| 7 dias | MAPE | LSTM | 0.01% |
| 7 dias | MSE | RF | 0.001 |
| 7 dias | MSE | SVR | 0.001 |
| 7 dias | MSE | LSTM | 0.001 |
| 7 dias | RMSE | RF | 0.037 |
| 7 dias | RMSE | SVR | 0.028 |
| 7 dias | RMSE | LSTM | 0.036 |
| 30 dias | MAE | RF | 0.036 |
| 30 dias | MAE | SVR | 0.033 |
| 30 dias | MAE | LSTM | 0.048 |
| 30 dias | MAPE | RF | 0.011% |
| 30 dias | MAPE | SVR | 0.01% |
| 30 dias | MAPE | LSTM | 0.014% |
| 30 dias | MSE | RF | 0.004 |
| 30 dias | MSE | SVR | 0.004 |
| 30 dias | MSE | LSTM | 0.005 |
| 30 dias | RMSE | RF | 0.066 |
| 30 dias | RMSE | SVR | 0.06 |
| 30 dias | RMSE | LSTM | 0.07 |
| 90 dias | MAE | RF | 0.039 |
| 90 dias | MAE | SVR | 0.039 |
| 90 dias | MAE | LSTM | 0.049 |
| 90 dias | MAPE | RF | 0.011% |
| 90 dias | MAPE | SVR | 0.011% |
| 90 dias | MAPE | LSTM | 0.014% |
| 90 dias | MSE | RF | 0.003 |
| 90 dias | MSE | SVR | 0.003 |
| 90 dias | MSE | LSTM | 0.004 |
| 90 dias | RMSE | RF | 0.059 |
| 90 dias | RMSE | SVR | 0.058 |
| 90 dias | RMSE | LSTM | 0.063 |
| 180 dias | MAE | RF | 0.05 |
| 180 dias | MAE | SVR | 0.059 |
| 180 dias | MAE | LSTM | 0.055 |
| 180 dias | MAPE | RF | 0.013% |
| 180 dias | MAPE | SVR | 0.016% |
| 180 dias | MAPE | LSTM | 0.015% |
| 180 dias | MSE | RF | 0.006 |
| 180 dias | MSE | SVR | 0.007 |
| 180 dias | MSE | LSTM | 0.006 |
| 180 dias | RMSE | RF | 0.075 |
| 180 dias | RMSE | SVR | 0.084 |
| 180 dias | RMSE | LSTM | 0.074 |

Fonte: Autoria própria.

Tabela 11 – Métricas de performance das previsões do produto Batata

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.071 |
| 7 dias | MAE | SVR | 0.041 |
| 7 dias | MAE | LSTM | 0.084 |
| 7 dias | MAPE | RF | 0.022% |
| 7 dias | MAPE | SVR | 0.013% |
| 7 dias | MAPE | LSTM | 0.027% |
| 7 dias | MSE | RF | 0.007 |
| 7 dias | MSE | SVR | 0.003 |
| 7 dias | MSE | LSTM | 0.008 |
| 7 dias | RMSE | RF | 0.083 |
| 7 dias | RMSE | SVR | 0.053 |
| 7 dias | RMSE | LSTM | 0.09 |
| 30 dias | MAE | RF | 0.072 |
| 30 dias | MAE | SVR | 0.047 |
| 30 dias | MAE | LSTM | 0.082 |
| 30 dias | MAPE | RF | 0.024% |
| 30 dias | MAPE | SVR | 0.015% |
| 30 dias | MAPE | LSTM | 0.027% |
| 30 dias | MSE | RF | 0.01 |
| 30 dias | MSE | SVR | 0.005 |
| 30 dias | MSE | LSTM | 0.01 |
| 30 dias | RMSE | RF | 0.098 |
| 30 dias | RMSE | SVR | 0.068 |
| 30 dias | RMSE | LSTM | 0.1 |
| 90 dias | MAE | RF | 0.087 |
| 90 dias | MAE | SVR | 0.071 |
| 90 dias | MAE | LSTM | 0.102 |
| 90 dias | MAPE | RF | 0.025% |
| 90 dias | MAPE | SVR | 0.02% |
| 90 dias | MAPE | LSTM | 0.029% |
| 90 dias | MSE | RF | 0.02 |
| 90 dias | MSE | SVR | 0.015 |
| 90 dias | MSE | LSTM | 0.026 |
| 90 dias | RMSE | RF | 0.141 |
| 90 dias | RMSE | SVR | 0.122 |
| 90 dias | RMSE | LSTM | 0.162 |
| 180 dias | MAE | RF | 0.123 |
| 180 dias | MAE | SVR | 0.147 |
| 180 dias | MAE | LSTM | 0.129 |
| 180 dias | MAPE | RF | 0.031% |
| 180 dias | MAPE | SVR | 0.034% |
| 180 dias | MAPE | LSTM | 0.033% |
| 180 dias | MSE | RF | 0.055 |
| 180 dias | MSE | SVR | 0.129 |
| 180 dias | MSE | LSTM | 0.057 |
| 180 dias | RMSE | RF | 0.235 |
| 180 dias | RMSE | SVR | 0.36 |
| 180 dias | RMSE | LSTM | 0.238 |

Fonte: Autoria própria.

Tabela 12 – Métricas de performance das previsões do produto Cebola

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.107 |
| 7 dias | MAE | SVR | 0.114 |
| 7 dias | MAE | LSTM | 0.161 |
| 7 dias | MAPE | RF | 0.03% |
| 7 dias | MAPE | SVR | 0.032% |
| 7 dias | MAPE | LSTM | 0.046% |
| 7 dias | MSE | RF | 0.013 |
| 7 dias | MSE | SVR | 0.015 |
| 7 dias | MSE | LSTM | 0.034 |
| 7 dias | RMSE | RF | 0.115 |
| 7 dias | RMSE | SVR | 0.123 |
| 7 dias | RMSE | LSTM | 0.184 |
| 30 dias | MAE | RF | 0.055 |
| 30 dias | MAE | SVR | 0.055 |
| 30 dias | MAE | LSTM | 0.09 |
| 30 dias | MAPE | RF | 0.015% |
| 30 dias | MAPE | SVR | 0.015% |
| 30 dias | MAPE | LSTM | 0.025% |
| 30 dias | MSE | RF | 0.005 |
| 30 dias | MSE | SVR | 0.005 |
| 30 dias | MSE | LSTM | 0.012 |
| 30 dias | RMSE | RF | 0.069 |
| 30 dias | RMSE | SVR | 0.071 |
| 30 dias | RMSE | LSTM | 0.108 |
| 90 dias | MAE | RF | 0.057 |
| 90 dias | MAE | SVR | 0.062 |
| 90 dias | MAE | LSTM | 0.07 |
| 90 dias | MAPE | RF | 0.015% |
| 90 dias | MAPE | SVR | 0.016% |
| 90 dias | MAPE | LSTM | 0.018% |
| 90 dias | MSE | RF | 0.006 |
| 90 dias | MSE | SVR | 0.006 |
| 90 dias | MSE | LSTM | 0.009 |
| 90 dias | RMSE | RF | 0.075 |
| 90 dias | RMSE | SVR | 0.077 |
| 90 dias | RMSE | LSTM | 0.095 |
| 180 dias | MAE | RF | 0.456 |
| 180 dias | MAE | SVR | 1.016 |
| 180 dias | MAE | LSTM | 0.122 |
| 180 dias | MAPE | RF | 0.065% |
| 180 dias | MAPE | SVR | 0.142% |
| 180 dias | MAPE | LSTM | 0.022% |
| 180 dias | MSE | RF | 0.967 |
| 180 dias | MSE | SVR | 4.846 |
| 180 dias | MSE | LSTM | 0.037 |
| 180 dias | RMSE | RF | 0.984 |
| 180 dias | RMSE | SVR | 2.201 |
| 180 dias | RMSE | LSTM | 0.193 |

Fonte: Autoria própria.

Tabela 13 – Métricas de performance das previsões do produto Cenoura

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.033 |
| 7 dias | MAE | SVR | 0.03 |
| 7 dias | MAE | LSTM | 0.037 |
| 7 dias | MAPE | RF | 0.012% |
| 7 dias | MAPE | SVR | 0.011% |
| 7 dias | MAPE | LSTM | 0.014% |
| 7 dias | MSE | RF | 0.001 |
| 7 dias | MSE | SVR | 0.001 |
| 7 dias | MSE | LSTM | 0.003 |
| 7 dias | RMSE | RF | 0.037 |
| 7 dias | RMSE | SVR | 0.038 |
| 7 dias | RMSE | LSTM | 0.054 |
| 30 dias | MAE | RF | 0.029 |
| 30 dias | MAE | SVR | 0.031 |
| 30 dias | MAE | LSTM | 0.032 |
| 30 dias | MAPE | RF | 0.012% |
| 30 dias | MAPE | SVR | 0.012% |
| 30 dias | MAPE | LSTM | 0.013% |
| 30 dias | MSE | RF | 0.001 |
| 30 dias | MSE | SVR | 0.001 |
| 30 dias | MSE | LSTM | 0.002 |
| 30 dias | RMSE | RF | 0.035 |
| 30 dias | RMSE | SVR | 0.038 |
| 30 dias | RMSE | LSTM | 0.044 |
| 90 dias | MAE | RF | 0.027 |
| 90 dias | MAE | SVR | 0.029 |
| 90 dias | MAE | LSTM | 0.035 |
| 90 dias | MAPE | RF | 0.012% |
| 90 dias | MAPE | SVR | 0.012% |
| 90 dias | MAPE | LSTM | 0.015% |
| 90 dias | MSE | RF | 0.001 |
| 90 dias | MSE | SVR | 0.001 |
| 90 dias | MSE | LSTM | 0.002 |
| 90 dias | RMSE | RF | 0.035 |
| 90 dias | RMSE | SVR | 0.038 |
| 90 dias | RMSE | LSTM | 0.044 |
| 180 dias | MAE | RF | 0.036 |
| 180 dias | MAE | SVR | 0.041 |
| 180 dias | MAE | LSTM | 0.042 |
| 180 dias | MAPE | RF | 0.019% |
| 180 dias | MAPE | SVR | 0.022% |
| 180 dias | MAPE | LSTM | 0.021% |
| 180 dias | MSE | RF | 0.003 |
| 180 dias | MSE | SVR | 0.005 |
| 180 dias | MSE | LSTM | 0.004 |
| 180 dias | RMSE | RF | 0.058 |
| 180 dias | RMSE | SVR | 0.07 |
| 180 dias | RMSE | LSTM | 0.064 |

Fonte: Autoria própria.

Tabela 14 – Métricas de performance das previsões do produto Laranja

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.016 |
| 7 dias | MAE | SVR | 0.027 |
| 7 dias | MAE | LSTM | 0.028 |
| 7 dias | MAPE | RF | 0.008% |
| 7 dias | MAPE | SVR | 0.012% |
| 7 dias | MAPE | LSTM | 0.013% |
| 7 dias | MSE | RF | 0.0 |
| 7 dias | MSE | SVR | 0.001 |
| 7 dias | MSE | LSTM | 0.001 |
| 7 dias | RMSE | RF | 0.017 |
| 7 dias | RMSE | SVR | 0.03 |
| 7 dias | RMSE | LSTM | 0.03 |
| 30 dias | MAE | RF | 0.027 |
| 30 dias | MAE | SVR | 0.038 |
| 30 dias | MAE | LSTM | 0.042 |
| 30 dias | MAPE | RF | 0.013% |
| 30 dias | MAPE | SVR | 0.018% |
| 30 dias | MAPE | LSTM | 0.02% |
| 30 dias | MSE | RF | 0.002 |
| 30 dias | MSE | SVR | 0.003 |
| 30 dias | MSE | LSTM | 0.003 |
| 30 dias | RMSE | RF | 0.04 |
| 30 dias | RMSE | SVR | 0.05 |
| 30 dias | RMSE | LSTM | 0.058 |
| 90 dias | MAE | RF | 0.022 |
| 90 dias | MAE | SVR | 0.026 |
| 90 dias | MAE | LSTM | 0.041 |
| 90 dias | MAPE | RF | 0.01% |
| 90 dias | MAPE | SVR | 0.012% |
| 90 dias | MAPE | LSTM | 0.019% |
| 90 dias | MSE | RF | 0.001 |
| 90 dias | MSE | SVR | 0.001 |
| 90 dias | MSE | LSTM | 0.003 |
| 90 dias | RMSE | RF | 0.032 |
| 90 dias | RMSE | SVR | 0.036 |
| 90 dias | RMSE | LSTM | 0.05 |
| 180 dias | MAE | RF | 0.02 |
| 180 dias | MAE | SVR | 0.026 |
| 180 dias | MAE | LSTM | 0.04 |
| 180 dias | MAPE | RF | 0.009% |
| 180 dias | MAPE | SVR | 0.012% |
| 180 dias | MAPE | LSTM | 0.019% |
| 180 dias | MSE | RF | 0.001 |
| 180 dias | MSE | SVR | 0.002 |
| 180 dias | MSE | LSTM | 0.003 |
| 180 dias | RMSE | RF | 0.032 |
| 180 dias | RMSE | SVR | 0.04 |
| 180 dias | RMSE | LSTM | 0.05 |

Fonte: Autoria própria.

Tabela 15 – Métricas de performance das previsões do produto Maçã

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.076 |
| 7 dias | MAE | SVR | 0.05 |
| 7 dias | MAE | LSTM | 0.091 |
| 7 dias | MAPE | RF | 0.011% |
| 7 dias | MAPE | SVR | 0.007% |
| 7 dias | MAPE | LSTM | 0.013% |
| 7 dias | MSE | RF | 0.007 |
| 7 dias | MSE | SVR | 0.006 |
| 7 dias | MSE | LSTM | 0.013 |
| 7 dias | RMSE | RF | 0.084 |
| 7 dias | RMSE | SVR | 0.077 |
| 7 dias | RMSE | LSTM | 0.113 |
| 30 dias | MAE | RF | 0.098 |
| 30 dias | MAE | SVR | 0.092 |
| 30 dias | MAE | LSTM | 0.131 |
| 30 dias | MAPE | RF | 0.014% |
| 30 dias | MAPE | SVR | 0.013% |
| 30 dias | MAPE | LSTM | 0.019% |
| 30 dias | MSE | RF | 0.016 |
| 30 dias | MSE | SVR | 0.014 |
| 30 dias | MSE | LSTM | 0.027 |
| 30 dias | RMSE | RF | 0.128 |
| 30 dias | RMSE | SVR | 0.119 |
| 30 dias | RMSE | LSTM | 0.163 |
| 90 dias | MAE | RF | 0.083 |
| 90 dias | MAE | SVR | 0.111 |
| 90 dias | MAE | LSTM | 0.113 |
| 90 dias | MAPE | RF | 0.011% |
| 90 dias | MAPE | SVR | 0.015% |
| 90 dias | MAPE | LSTM | 0.015% |
| 90 dias | MSE | RF | 0.012 |
| 90 dias | MSE | SVR | 0.018 |
| 90 dias | MSE | LSTM | 0.018 |
| 90 dias | RMSE | RF | 0.108 |
| 90 dias | RMSE | SVR | 0.134 |
| 90 dias | RMSE | LSTM | 0.133 |
| 180 dias | MAE | RF | 0.085 |
| 180 dias | MAE | SVR | 0.105 |
| 180 dias | MAE | LSTM | 0.098 |
| 180 dias | MAPE | RF | 0.011% |
| 180 dias | MAPE | SVR | 0.014% |
| 180 dias | MAPE | LSTM | 0.013% |
| 180 dias | MSE | RF | 0.012 |
| 180 dias | MSE | SVR | 0.016 |
| 180 dias | MSE | LSTM | 0.015 |
| 180 dias | RMSE | RF | 0.112 |
| 180 dias | RMSE | SVR | 0.126 |
| 180 dias | RMSE | LSTM | 0.122 |

Fonte: Autoria própria.

Tabela 16 – Métricas de performance das previsões do produto Mamão

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.091 |
| 7 dias | MAE | SVR | 0.044 |
| 7 dias | MAE | LSTM | 0.168 |
| 7 dias | MAPE | RF | 0.017% |
| 7 dias | MAPE | SVR | 0.008% |
| 7 dias | MAPE | LSTM | 0.031% |
| 7 dias | MSE | RF | 0.012 |
| 7 dias | MSE | SVR | 0.003 |
| 7 dias | MSE | LSTM | 0.034 |
| 7 dias | RMSE | RF | 0.11 |
| 7 dias | RMSE | SVR | 0.05 |
| 7 dias | RMSE | LSTM | 0.183 |
| 30 dias | MAE | RF | 0.098 |
| 30 dias | MAE | SVR | 0.075 |
| 30 dias | MAE | LSTM | 0.121 |
| 30 dias | MAPE | RF | 0.024% |
| 30 dias | MAPE | SVR | 0.019% |
| 30 dias | MAPE | LSTM | 0.029% |
| 30 dias | MSE | RF | 0.014 |
| 30 dias | MSE | SVR | 0.008 |
| 30 dias | MSE | LSTM | 0.024 |
| 30 dias | RMSE | RF | 0.119 |
| 30 dias | RMSE | SVR | 0.092 |
| 30 dias | RMSE | LSTM | 0.154 |
| 90 dias | MAE | RF | 0.084 |
| 90 dias | MAE | SVR | 0.063 |
| 90 dias | MAE | LSTM | 0.106 |
| 90 dias | MAPE | RF | 0.023% |
| 90 dias | MAPE | SVR | 0.018% |
| 90 dias | MAPE | LSTM | 0.029% |
| 90 dias | MSE | RF | 0.017 |
| 90 dias | MSE | SVR | 0.012 |
| 90 dias | MSE | LSTM | 0.027 |
| 90 dias | RMSE | RF | 0.13 |
| 90 dias | RMSE | SVR | 0.11 |
| 90 dias | RMSE | LSTM | 0.163 |
| 180 dias | MAE | RF | 0.096 |
| 180 dias | MAE | SVR | 0.077 |
| 180 dias | MAE | LSTM | 0.134 |
| 180 dias | MAPE | RF | 0.025% |
| 180 dias | MAPE | SVR | 0.02% |
| 180 dias | MAPE | LSTM | 0.035% |
| 180 dias | MSE | RF | 0.019 |
| 180 dias | MSE | SVR | 0.015 |
| 180 dias | MSE | LSTM | 0.035 |
| 180 dias | RMSE | RF | 0.138 |
| 180 dias | RMSE | SVR | 0.122 |
| 180 dias | RMSE | LSTM | 0.186 |

Fonte: Autoria própria.

Tabela 17 – Métricas de performance das previsões do produto Melancia

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.041 |
| 7 dias | MAE | SVR | 0.014 |
| 7 dias | MAE | LSTM | 0.051 |
| 7 dias | MAPE | RF | 0.015% |
| 7 dias | MAPE | SVR | 0.005% |
| 7 dias | MAPE | LSTM | 0.018% |
| 7 dias | MSE | RF | 0.002 |
| 7 dias | MSE | SVR | 0.0 |
| 7 dias | MSE | LSTM | 0.003 |
| 7 dias | RMSE | RF | 0.045 |
| 7 dias | RMSE | SVR | 0.018 |
| 7 dias | RMSE | LSTM | 0.054 |
| 30 dias | MAE | RF | 0.053 |
| 30 dias | MAE | SVR | 0.03 |
| 30 dias | MAE | LSTM | 0.048 |
| 30 dias | MAPE | RF | 0.018% |
| 30 dias | MAPE | SVR | 0.011% |
| 30 dias | MAPE | LSTM | 0.017% |
| 30 dias | MSE | RF | 0.004 |
| 30 dias | MSE | SVR | 0.001 |
| 30 dias | MSE | LSTM | 0.003 |
| 30 dias | RMSE | RF | 0.061 |
| 30 dias | RMSE | SVR | 0.035 |
| 30 dias | RMSE | LSTM | 0.057 |
| 90 dias | MAE | RF | 0.048 |
| 90 dias | MAE | SVR | 0.037 |
| 90 dias | MAE | LSTM | 0.041 |
| 90 dias | MAPE | RF | 0.023% |
| 90 dias | MAPE | SVR | 0.019% |
| 90 dias | MAPE | LSTM | 0.02% |
| 90 dias | MSE | RF | 0.004 |
| 90 dias | MSE | SVR | 0.003 |
| 90 dias | MSE | LSTM | 0.003 |
| 90 dias | RMSE | RF | 0.062 |
| 90 dias | RMSE | SVR | 0.051 |
| 90 dias | RMSE | LSTM | 0.055 |
| 180 dias | MAE | RF | 0.044 |
| 180 dias | MAE | SVR | 0.042 |
| 180 dias | MAE | LSTM | 0.041 |
| 180 dias | MAPE | RF | 0.022% |
| 180 dias | MAPE | SVR | 0.022% |
| 180 dias | MAPE | LSTM | 0.021% |
| 180 dias | MSE | RF | 0.004 |
| 180 dias | MSE | SVR | 0.004 |
| 180 dias | MSE | LSTM | 0.003 |
| 180 dias | RMSE | RF | 0.063 |
| 180 dias | RMSE | SVR | 0.063 |
| 180 dias | RMSE | LSTM | 0.059 |

Fonte: Autoria própria.

Tabela 18 – Métricas de performance das previsões do produto Tomate

| Período | Métrica | Modelo | Valor |
|----------|---------|--------|--------|
| 7 dias | MAE | RF | 0.078 |
| 7 dias | MAE | SVR | 0.06 |
| 7 dias | MAE | LSTM | 0.059 |
| 7 dias | MAPE | RF | 0.021% |
| 7 dias | MAPE | SVR | 0.016% |
| 7 dias | MAPE | LSTM | 0.016% |
| 7 dias | MSE | RF | 0.01 |
| 7 dias | MSE | SVR | 0.005 |
| 7 dias | MSE | LSTM | 0.005 |
| 7 dias | RMSE | RF | 0.098 |
| 7 dias | RMSE | SVR | 0.072 |
| 7 dias | RMSE | LSTM | 0.072 |
| 30 dias | MAE | RF | 0.086 |
| 30 dias | MAE | SVR | 0.07 |
| 30 dias | MAE | LSTM | 0.092 |
| 30 dias | MAPE | RF | 0.026% |
| 30 dias | MAPE | SVR | 0.022% |
| 30 dias | MAPE | LSTM | 0.029% |
| 30 dias | MSE | RF | 0.016 |
| 30 dias | MSE | SVR | 0.011 |
| 30 dias | MSE | LSTM | 0.017 |
| 30 dias | RMSE | RF | 0.127 |
| 30 dias | RMSE | SVR | 0.102 |
| 30 dias | RMSE | LSTM | 0.132 |
| 90 dias | MAE | RF | 0.094 |
| 90 dias | MAE | SVR | 0.082 |
| 90 dias | MAE | LSTM | 0.112 |
| 90 dias | MAPE | RF | 0.028% |
| 90 dias | MAPE | SVR | 0.025% |
| 90 dias | MAPE | LSTM | 0.035% |
| 90 dias | MSE | RF | 0.02 |
| 90 dias | MSE | SVR | 0.015 |
| 90 dias | MSE | LSTM | 0.023 |
| 90 dias | RMSE | RF | 0.142 |
| 90 dias | RMSE | SVR | 0.124 |
| 90 dias | RMSE | LSTM | 0.151 |
| 180 dias | MAE | RF | 0.121 |
| 180 dias | MAE | SVR | 0.121 |
| 180 dias | MAE | LSTM | 0.146 |
| 180 dias | MAPE | RF | 0.031% |
| 180 dias | MAPE | SVR | 0.031% |
| 180 dias | MAPE | LSTM | 0.038% |
| 180 dias | MSE | RF | 0.033 |
| 180 dias | MSE | SVR | 0.036 |
| 180 dias | MSE | LSTM | 0.04 |
| 180 dias | RMSE | RF | 0.183 |
| 180 dias | RMSE | SVR | 0.189 |
| 180 dias | RMSE | LSTM | 0.2 |

Fonte: Autoria própria.