**UNIVERSIDADE FEDERAL DA PARAÍBA**

**CENTRO DE CIÊNCIAS HUMANAS, SOCIAIS E AGRÁRIAS**

**DEPARTAMENTO DE CIÊNCIAS SOCIAIS APLICADAS**

**CURSO DE BACHARELADO EM ADMINISTRAÇÃO**

LUCAS SOUZA SANTOS

**APLICAÇÃO DE MODELO DE REGRESSÃO DE VETOR POR SUPORTE NA PREVISÃO DE PREÇOS DE COMMODITIES:** UM CASO DA SOJA

Bananeiras - PB

2023LUCAS SOUZA SANTOS

**APLICAÇÃO DE MODELO DE REGRESSÃO DE VETOR POR SUPORTE NA PREVISÃO DE PREÇOS DE COMMODITIES:** UM CASO DA SOJA

Projeto de trabalho de conclusão de curso apresentado como parte do requisito à obtenção do grau de Bacharel, pelo curso de Graduação em Administração do Departamento de Ciências Sociais Aplicadas da Universidade Federal da Paraíba.

Orientador(a): Prof. O Dr. Gustavo Correa Xavier

Bananeiras - PB

2023

**SUMÁRIO**

[**1 INTRODUÇÃO 4**](#_heading=h.gjdgxs)

[1.1 OBJETIVOS 5](#_heading=h.35nkun2)

[**1.1.1 Objetivo geral 5**](#_heading=h.1ksv4uv)

[**1.1.2 Objetivos específicos 5**](#_heading=h.44sinio)

[**2 REFERENCIAL TEÓRICO 6**](#_heading=h.2jxsxqh)

[2.1 IMPACTO DAS COMMODITIES NO MERCADO BRASILEIRO 6](#_heading=h.z337ya)

[2.2 PREVISÃO DE VARIÁVEIS EM SÉRIES TEMPORAIS 8](#_heading=h.3j2qqm3)

[**3 MÉTODO DA PESQUISA 9**](#_heading=h.1y810tw)

[**REFERÊNCIAS 12**](#_heading=)

[**APÊNDICE A – Cronograma de Execução do Projeto 14**](#_heading=h.2xcytpi)

1. **INTRODUÇÃO**

O mercado financeiro é enormemente vasto em produtos de investimento e instrumentos através dos quais se pode investir, é comum deparar-nos com instrumentos para investir como Ações, ETFs - Fundos de Índices, Derivativos, dentre outros. E um dos produtos financeiros que são de grande importância para as economias e no qual pode-se investir são as commodities (plural de commodity). Uma commodity pode ser grãos, ouro, carne bovina, petróleo, gás natural, etc. Contanto que possa ser definida como um ativo físico que possui características padronizadas, de ampla negociação em diversas localidades e que possa ser transportado e armazenado por um longo período sem perda da qualidade, também sendo conhecidos por insumos ou matérias-primas (MOLERO; MELLO, 2021).

Molero e Mello (2021) afirmam ainda que há tipos de mercados diferentes em que as commodities podem ser negociadas, a primeira é no mercado à vista, que envolve entrega física e imediata da mercadoria, a segunda é ser negociada como referência nos mercados derivativos. Segundo Giambiagi (2017) os derivativos são instrumentos financeiros cujo resultado depende da realização de uma variável aleatória, normalmente representado por um preço de um ativo em uma data futura. Eles existem para permitir a transferência de risco entre agentes da economia. Os tipos de derivativos mais comuns são: termo, opção, swap e futuro, sendo um dos mais comuns.

Ao longo das últimas décadas vários autores se dedicaram ao estudo e avaliação de produtos financeiros, tal qual as commodities. Sørensen (2002) por exemplo, modelou séries temporais de commodities agrícolas para estimar preços futuros de cereais como milho, soja e trigo. Richter e Sørensen (2002) estimaram um modelo de volatilidade estocástica de dados de Futuros e Opções da soja. Lima et al. (2009) obteve resultados satisfatórios ao realizar estudos baseados na realização de previsões dentro das subséries decompostas de preços da soja através de Wavelets (onduletas) em conjunto com modelos econométricos e comparando estes com previsões por modelos de Redes Neurais.

Entender e estimar o comportamento dos preços de commodities mostra-se um atraente e relevante campo de estudo e de geração de análises para produtores e investidores. Estes preços variam no tempo, ao passo que suas negociações acontecem no mercado, a essa variação do preço no tempo podemos chamar de *Times Series* ou Séries Temporais. As séries temporais sejam, a temperatura de uma cidade ao longo de um ano, quantidade de vendas em um semestre, ou o preço do índice Ibovespa ao longo de sua existência, são objetos valiosos para extração de insights que ajudem seus stakeholders (partes interessadas) na sua tomada de decisão.

Várias tecnologias, instrumentos matemáticos e modelos econométricos têm sido aplicados na análise de séries temporais financeiras, por serem ferramentas interdisciplinares. Uma dessas técnicas mais recentes são, Máquinas de Vetor de Suporte (SVM) ou Support Vector Machine do inglês, que é uma técnica de Aprendizado de Máquina (AM) aplicada em tarefas de aprendizado supervisionado (KIM, 2003; LIMA et al, 2010; LAHMIRI, 2013; TAY e CAO, 2001).

Como também as já conhecidas técnicas econométricas aplicadas por autores como Lima et al. (2010) que utilizou modelos ARIMA-GARCH, na modelagem de tendência, sazonalidade e volatilidade. Pai e Lin (2005) que utilizaram um modelo híbrido ARIMA e Máquinas Vetoriais de Suporte na previsão de preços de ações. Assim o objetivo deste estudo é compreender descritivamente a produção da commodity brasileira da soja para entender como essas características se relacionam as séries de preço, em seguida aplicar as técnicas de decomposição de séries temporais em preços da soja e modelos de previsão através do SVR (Regressão Vetorial de Suporte) um sub-modelo do SVM, explorando diferentes kernels e parâmetros. Para assim compreender e avaliar a validade do uso destas técnicas na previsão de variáveis temporais financeiras de commodities.

Este trabalho encontra-se dividido além desta parte (1) introdutória, em (2) referencial teórico, (3) método da pesquisa, (4) resultados e (5) conclusões.

* 1. OBJETIVOS
     1. **Objetivo geral**

Explorar a viabilidade da aplicação da metodologia de decomposição e previsão de séries temporais de preços de commodities, por meio de modelos de regressão de Máquinas de Vetor de Suporte (SVM).

* + 1. **Objetivos específicos**

1. Realizar a análise descritiva da produção da soja no Brasil.
2. Realizar uma análise detalhada da aplicação da decomposição das séries temporais de preços da soja, identificando os componentes de tendência, sazonalidade e ruído.
3. Implementar o modelo Regressão Vetorial de Suporte (SVR) para a previsão dos preços da soja, explorando diferentes kernels e parâmetros.
4. Avaliar o desempenho do modelo de previsão, considerando métricas de precisão, estabilidade e robustez, a fim de verificar a viabilidade prática dessa abordagem na previsão de séries temporais financeiras.
5. **REFERENCIAL TEÓRICO**
   1. IMPACTO DAS COMMODITIES NO MERCADO BRASILEIRO

As commodities tiveram forte influência no crescimento econômico brasileiro, durante o famoso boom das commodities que perdurou dos anos 2000-2014. E recentemente, porém, o crescimento desacelerou, em parte como consequência do declínio do preço das commodities (BLANCHARD, 2017). Além do Brasil, toda a América Latina e Caribe (ALC) se beneficiou da formação de capital e crescimento das exportações nesse período. Em seu estudo Kristjanpoller, Olson e Salazar (2016) mostram que as commodities no geral tiveram efeito positivo no crescimento da ALC durante o período de boom, mais especificamente os combustíveis e os produtos industriais.

Mostrando a força da indústria extrativa e agrícola brasileira, as commodities representam hoje cerca de 70% das exportações brasileiras e cresceram 16,8% em volume na comparação entre julho de 2022 e 2023; e 9,8% na comparação interanual do primeiro semestre desses anos. Os preços recuaram em 17% na comparação mensal, e 8,6% na comparação entre os acumulados do ano até julho. O volume exportado das não commodities registrou queda de 1,1% entre os meses de julho de 2022 e 2023, e cresceu 0,6% na comparação entre os acumulados do ano até julho (FGV IBRE, 2023, p.23).

Compreende-se então a relevância que as commodities têm para a economia mundial, se tornando assim nas últimas décadas base de vários produtos no mercado financeiro em todo o mundo. Na análise entre esses mercados com o mercado acionário, Brooks e Prokopczuk (2013) concluem que as commodities podem ser um diversificador útil da volatilidade e dos retornos das ações, podendo ser um recurso útil para a constituição de portfólio.

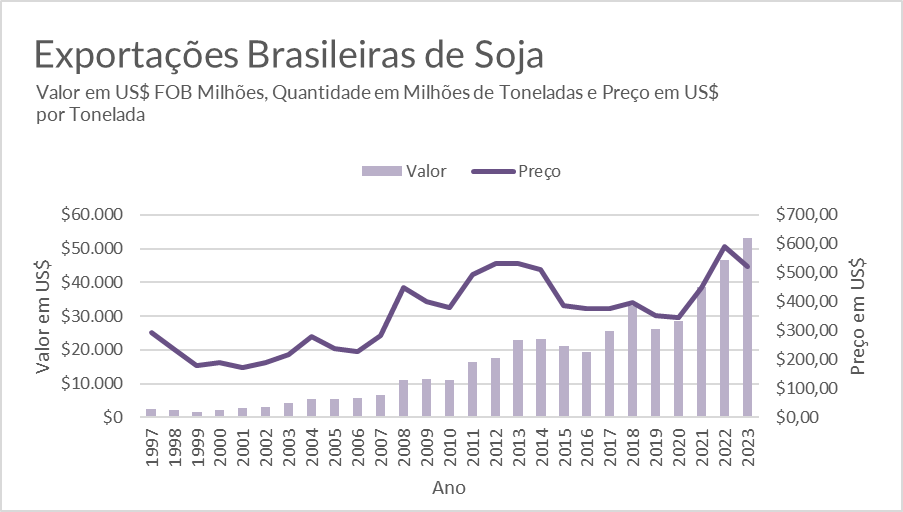
Uma das áreas de crescimento mais rápido nas finanças empíricas, e uma das menos rigorosamente analisadas, especialmente do ponto de vista da econometria financeira, é a análise econométrica de derivados financeiros. Juntamente com as ações e as obrigações, os derivados constituem a terceira categoria principal de instrumentos financeiros e são normalmente negociados em bolsa ou no mercado de balcão (CHANG; MCALEER, 2015, p.1)

Haase, Zimmermann e Zimmermann (2016) revisaram estudos empíricos sobre o impacto da especulação no Mercado Futuro de Commodities (MFC), e concluem que não há consenso entre os economistas sobre o impacto da especulação no MFC, porém a especulação de preços nesse mercado para commodities agrícolas é vista como uma preocupação, pois pode levar a aumentos nos preços dos alimentos, afetando negativamente as economias menos desenvolvidas.

Assim, é importante para investidores, produtores, consumidores e demais tomadores de decisão ter uma boa compreensão do comportamento dos preços das commodities e de suas interdependências, bem como da sua relação com o mercado acionista (BROOKS e PROKOPCZUK, 2013, p.1).

Como visto anteriormente, existem várias commodities, dentre as que mais se destacam no Brasil estão as agrícolas, principalmente a soja que segundo os dados consolidados de 2023 da Secretaria do Comércio Exterior do Brasil (2023) é a mais exportada do país estando em 1º no ranking.

**Gráfico 01:** Exportações brasileiras de soja.



**Fonte:** Elaboração própria a partir dos dados de (BRASIL, 2023).

O “Gráfico 01” acima apresenta a evolução do valor da soja brasileira exportada e o preço da tonelada de 1997 a 2023, no consolidado de 2023 ela atingiu seu recorde de US$ 53.244,60 milhões em soja exportada, a um preço de US$ 522,70 a tonelada.

Tendo em vista a relevância da soja não só para o Brasil, a soja já vem sendo abordada em alguns estudos financeiros sobre commodities, como em Richter e Sørensen (2002) que abordaram um modelo de volatilidade estocástica em tempo contínuo usando dados em painel de Futuros e Opções de soja. Brooks e Prokopczuk (2013) que estudaram o comportamento estocástico dos preços e volatilidades de uma amostra de seis dos mais importantes mercados de commodities (inclusive a soja) e compararam essas propriedades com as do mercado acionário, usando Cadeias de Markov Monte Carlo (CMMC). Além de estudos sobre previsão de preços de commodities agrícolas usando estrutura de seleção de modelo com séries temporais, analisando o uso de modelos de Redes Neurais Artificiais (RNA), Máquina de Aprendizagem Extrema (MAE) e o modelo que enfatizamos neste estudo, Máquinas de Vetor de Suporte (SVM) (ZHANG et al, 2020).

* 1. PREVISÃO DE VARIÁVEIS EM SÉRIES TEMPORAIS

Segundo Petropoulos et al. (2022) apesar da vasta literatura econométrica e de previsão de modelos de fatores comuns, há poucos trabalhos até agora sobre a aplicação de tais modelos para commodities agrícolas e suas séries temporais, além de vários desses trabalhos serem recentes, o que indica que existem muitos caminhos abertos para investigação futura sobre estes tópicos e, em particular, para a previsão aplicada.

Os rápidos avanços na computação permitiram a análise de conjuntos de dados maiores e mais complexos e estimularam o interesse em análise e ciência de dados. Como resultado, a caixa de ferramentas de métodos dos previsores cresceu em tamanho e sofisticação. A ciência da computação abriu o caminho com métodos como redes neurais e outros tipos de aprendizado de máquina, que estão recebendo muita atenção de analistas e tomadores de decisão (PETROPOULOS et al, 2022, p.7).

Porém, como Haase, Zimmermann e Zimmermann (2016) ressaltam, a previsão de preços de commodities agrícolas é um desafio por vezes complexo devido sua natureza volátil e à influência de múltiplos fatores, dependendo da commodity analisada. Sendo assim é necessária uma variedade de informações e modelos para obter previsões mais precisas. Para a previsão aplicada às commodities alguns autores como no caso de Zhang et al. (2020) têm proposto uma estrutura de seleção de modelo para previsão de séries temporais, baseados em séries temporais características e horizontes de previsão.

Séries temporais financeiras podem gerar vários insights e possibilidades de análise, e no uso financeiro da previsão de preços, como no trabalho de Lahmiri (2013) que faz previsões da tendência futura do índice de preços do S&P500 por meio de séries temporais de baixa frequência extraídas de uma análise Wavelet.

Existem vários métodos de previsão na econometria e no aprendizado de máquina que têm sido amplamente encontrados na literatura econômica e financeira. Por exemplo, Gujarati, Yamagama e Girvilitto (2019) apresentam alguns dos principais, os quais são (1) modelos de regressão, (2) o processo autorregressivo integrado de médias móveis (ARIMA), popularizado pelos estatísticos Box e Jenkins, também conhecido como metodologia Box-Jenkins (BJ) e (3) vetores autorregressivos (VAR).

Entretanto, Petropoulos et al. (2022) apresentam outros modelos, dentre eles o modelo de Regressão Vetorial de Suporte (SVR) para problemas de previsão, como bastante úteis para previsão de consumo de energia e previsão de preços de petróleo, por forneceram soluções poderosas para reconhecer padrões não lineares e irregulares. O modelo SVR é uma versão do SVM, que foi proposto por Drucker et al (1996).

**Figura 01:** Hiperplano de separação ótima e seus hiperplanos de suporte. Os eixos ordenados x1 e x2 representam as dimensões das amostras no espaço 2D.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

**Fonte:** (CAMPOS, 2020).

Conforme a “Figura 01” apresentada acima, é colocado por Campos (2020, p.23) que o conceito de generalização das SVM pode ser exemplificado com uma classificação binária. A partir de duas classes e um conjunto de dados, as SVM determinam o hiperplano que os separa, de maneira a colocar a maior quantidade possível de pontos da mesma classe do mesmo lado.

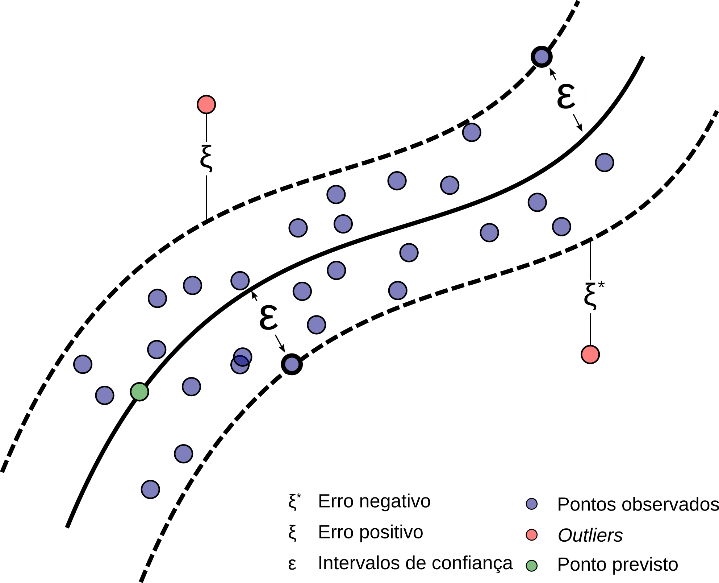
1. **MÉTODO DA PESQUISA**

A metodologia adotada neste estudo se baseia na aplicação da decomposição de séries temporais e no modelo de regressão SVM (Support Vector Machine) de aprendizado não supervisionado, com o objetivo de analisar séries temporais não lineares de preços da commodity soja, considerando sacas de 60kg conforme dados fornecidos pela CEPEA/ESALQ/USP. Além disso, realizamos uma análise inicial para caracterizar a produção de soja no Brasil.

A abordagem metodológica deste estudo pode ser categorizada como quantitativa e descritiva. Quanto aos objetivos, esta pesquisa se concentra na descrição e análise dos dados relacionados à série temporal de preços da soja e à sua produção no contexto brasileiro, o que a caracteriza como descritiva. Além disso, a abordagem é quantitativa, uma vez que faz uso de métodos estatísticos e de Machine Learning para análise dos dados. Os passos metodológicos podem ser resumidos da seguinte forma:

1. Preparação dos dados, com a obtenção dos dados da soja do site da CEPEA/ESALQ/USP além de outros dados sobre a commodity obtidos nos sites do IBGE (Instituto Brasileiro de Geografia e Estatística) e do SIDRA, e o carregamento da série temporal de preços da soja no algoritmo.
2. Análise descritiva dos dados da commodity.
3. Decomposição da série para análise de sazonalidade, tendência e ruído.
4. Preparação da série de preços para a regressão SVM.
5. Divisão dos dados em conjuntos de treinamento e teste.
6. Treino do modelo SVM de regressão. Escolha de um kernel SVM adequado (por exemplo: 'linear', 'rbf', ‘poly’ etc).
7. Previsões e avaliação do desempenho do modelo. Uso do modelo treinado para fazer previsões no conjunto de testes. Avaliação do desempenho do modelo usando métricas relevantes, como o erro quadrático médio (MSE) ou outras métricas apropriadas.

**Figura 02:** Representação de hiperplano não linear para regressão.



**Fonte:** (ALMEIDA; CARVALHO; MENINO, 2017).

No SVR conforme a “Figura: 02” os valores são previstos através dos hiperplanos, que utilizam derivadas parciais para os cálculos dos intervalos de confiança. As margens (ϵ) representam os intervalos de confiança e os vetores de suporte que as delimitam, representam os limites para os erros positivos () e negativos () (ALMEIDA; CARVALHO; MENINO, 2017).

**REFERÊNCIAS**

ALMEIDA, A; CARVALHO, F; MENINO, F. **Introdução ao Machine Learning:** de alunos para alunos. [S. L.]: Dataat, 2017.

BLANCHARD, Olivier. **Macroeconomia**. 7. ed. São Paulo, SP: Pearson, 2017. E-book. Disponível em: https://plataforma.bvirtual.com.br. Acesso em: 19 ago. 2023.

BRASIL. Mdic. Secretaria de Comércio Exterior. **Comex Vis**. Disponível em: http://comexstat.mdic.gov.br/pt/comex-vis. Acesso em: 01 set. 2023.

BROOKS, C; PROKOPCZUK, M. The dynamics of commodity prices. **Quantitative Finance**, [S.L.], v. 13, n. 4, p. 527-542, abr. 2013. Informa UK Limited. <http://dx.doi.org/10.1080/14697688.2013.769689>.

CAMPOS, B. A. R de. M. **Análise Comparativa de Técnicas para a Previsão de Séries Temporais no Contexto de Mercado Financeiros**. 2020. 76 f. TCC (Graduação) - Curso de Sistemas de Informação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis - Sc, 2021.

CEPEA/ESALQ/USP – Centro de Estudos Avançados em Economia Aplicada / Escola Superior de Agricultura Luiz de Queiroz / Universidade de São Paulo. (2023). <https://www.cepea.esalq.usp.br/br>.

CHANG, C; MCALEER, M. Econometric analysis of financial derivatives: an overview. **Journal Of Econometrics**, [S.L.], v. 187, n. 2, p. 403-407, ago. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.jeconom.2015.02.026>.

DRUCKER, H. et al. **Support Vector Regression Machines**. Em: NIPS. 3 dez. 1996. Disponível em: <https://www.semanticscholar.org/paper/Support-Vector-Regression-Machines-Drucker-Burges/e52fb14e4beccc5e88a33c1fe5c7d6e780831ae1>. Acesso em: 4 set. 2023.

FGV IBRE. **Boletim Macro**. 2023. Disponível em: https://portalibre.fgv.br/sites/default/files/2023-08/2023%2008%20Boletim%20Macro\_1.pdf. Acesso em: 27 ago. 2023.

GIAMBIAGI, F. **Derivativos e Risco de Mercado**. Rio de Janeiro - RJ: Grupo GEN, 2017. E-book. ISBN 9788595154742.

GUJARATI, D; YAMAGAMI, C; VIRGILITTO, S. B. **Econometria**. São Paulo - SP: Editora Saraiva, 2019. E-book. ISBN 9788553131952.

HAASE, M; ZIMMERMANN, Y. S; ZIMMERMANN, H. The impact of speculation on commodity futures markets – A review of the findings of 100 empirical studies. **Journal Of Commodity Markets**, [S.L.], v. 3, n. 1, p. 1-15, set. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.jcomm.2016.07.006.

KIM, K. Financial time series forecasting using support vector machines. **Neurocomputing**, [S.L.] v. 55, n. 1–2, p. 307–319, set. 2003. Elsevier BV. <http://dx.doi.org/10.1016/s0925-2312(03)00372-2>.

KRISTJANPOLLER, W; OLSON, J. E; SALAZAR, R. I. Does the commodities boom support the export led growth hypothesis? Evidence from Latin American countries. **Latin American Economic Review**, [S.L.], v. 25, n. 1, 20 out. 2016. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s40503-016-0036-z.

LAHMIRI, S. Forecasting Direction of the S&P 500 Movement Using Wavelet Transform and Support Vector Machines. **International Journal of Strategic Decision Sciences**, [S.L.], v. 4, n. 1, p. 79–89, 1 jan. 2013. IGI Global. <http://dx.doi.org/10.4018/jsds.2013010105>.

LIMA, F. G. et al. Previsão de preços de commodities com modelos ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias - novos resultados. **Revista de Administração**, [S.L.], v. 45, n. 2, p. 188-202, abr. 2010. Elsevier BV. http://dx.doi.org/10.1016/s0080-2107(16)30537-4.

MOLERO, L; MELLO, E. M. **Derivativos – Negociação e precificação 2ª edição**. São Paulo - SP: Saint Paul Publishing (Brazil), 2021. E-book. ISBN 9786586407150.

PAI, P; LIN, C. A hybrid ARIMA and support vector machines model in stock price forecasting. **Omega**, [S.L.], v. 33, n. 6, p. 497-505, dez. 2005. Elsevier BV. <http://dx.doi.org/10.1016/j.omega.2004.07.024>.

PETROPOULOS, F. et al. Forecasting: theory and practice. **International Journal of Forecasting**, [S.L.], v. 38, n. 3, p. 705-871, jul. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.ijforecast.2021.11.001.

RICHTER, M. C.; SØRENSEN, C. Stochastic Volatility and Seasonality in Commodity Futures and Options: The Case of Soybeans. **SSRN Electronic Journal**, [S.L.], 2002. Elsevier BV. <http://dx.doi.org/10.2139/ssrn.301994>.

SØRENSEN, C. Modeling seasonality in agricultural commodity futures. **Journal Of Futures Markets**, [S.L.], v. 22, n. 5, p. 393-426, 7 mar. 2002. Wiley. <http://dx.doi.org/10.1002/fut.10017>.

TAY, F. E. H; CAO, L. Application of support vector machines in financial time series forecasting. **Omega**, [S.L.], v. 29, n. 4, p. 309-317, ago. 2001. Elsevier BV. <http://dx.doi.org/10.1016/s0305-0483(01)00026-3>.

ZHANG, D. et al. Forecasting Agricultural Commodity Prices Using Model Selection Framework with Time Series Features and Forecast Horizons. **Ieee Access**, [S.L.], v. 8, p. 28197-28209, 2020. Institute of Electrical and Electronics Engineers (IEEE). http://dx.doi.org/10.1109/access.2020.2971591.

1. **– Cronograma de Execução do Projeto**



|  | **Data** | **2023** | | | | | | **2024** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Etapa** |  | **Jul.** | **Ago.** | **Set.** | **Out.** | **Nov.** | **Dez.** | **Jan.** | **Fev.** | **Mar.** | **Abr.** | **Maio** |
| Etapa 1: Definição do problema de pesquisa e Orientador. | | X | X |  |  |  |  |  |  |  |  |  |
| Etapa 2: Revisão bibliográfica. | |  | X | X | X |  |  |  |  |  |  |  |
| Etapa 3: Escrita e entrega do TCC 1 | |  | X | X | X | X |  |  |  |  |  |  |
| Etapa 4: Codificação do modelo. | |  |  |  |  | X |  |  |  |  |  |  |
| Etapa 5: Coleta dos dados. | |  |  |  |  | X | X | X |  |  |  |  |
| Etapa 6: Análise dos dados. | |  |  |  |  | X | X | X | X | X | X | X |