

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA E DE ARTES
ENGENHARIA DE CONTROLE E AUTOMAÇÃO
Trabalho Final de Curso II

Felipe Augusto Romano
Matheus Fernandes Rodrigues de Sousa

**ANÁLISE PREDITIVA DA PRECIFICAÇÃO DE COMMODITIES PECUÁRIAS
UTILIZANDO REDES NEURAIS LSTM**

Trabalho Final de Curso II como parte dos requisitos para
obtenção do título de bacharel em Engenharia de
Controle e Automação apresentado à Pontifícia
Universidade Católica de Goiás.

BANCA EXAMINADORA:

Prof. Dr. Marcos Antônio de Sousa – Orientador. POLI-PUC Goiás.
Prof. Dr. Antônio Marcos de Melo Medeiros – POLI-PUC Goiás.
Prof. Dr. Bruno Quirino de Oliveira – POLI-PUC Goiás.

Goiânia, 14 de junho de 2023.

Análise Preditiva da Precificação de Commodities Pecuárias Utilizando Redes Neurais LSTM

FELIPE AUGUSTO ROMANO, MATHEUS FERNANDES R. DE SOUSA, MARCOS A. SOUSA

Abstract – This undergraduate thesis presents the development of an Artificial Neural Network (ANN) based on Time Series to make predictions on livestock commodity values, with a focus on beef. The system is capable of analyzing input data and generating graphs that show the expected predictions for the future values of these commodities, considering the prediction analysis of the price of the next day based on the previous ten days. Scenarios were carried out using time series with price data for a fat cow's arroba between 1997 and 2022, which were analyzed and discussed. The preliminary results of training and prediction demonstrated the effectiveness of the developed computational tool, regarding the accuracy of the implemented prediction model. Thus, the applicability of the LSTM Artificial Neural Network as an effective technique for predicting livestock commodity values, mainly beef, can be observed.

Keywords – Artificial neural network, artificial intelligence, prediction, farming, data analysis, time series, recurrent neural networks.

Resumo – Este trabalho de conclusão de curso apresenta o desenvolvimento de uma Rede Neural Artificial (RNA) baseada em Séries Temporais para realizar previsões nos valores de commodities pecuárias, com destaque para a carne bovina. O sistema é capaz de analisar dados de entrada e gerar gráficos que mostram as previsões esperadas para os valores futuros dessas commodities, considerando a análise de predição do preço do dia seguinte com base nos dez dias anteriores. Foram realizados cenários utilizando séries temporais com dados de preço da arroba do boi gordo, entre os anos de 1997 e 2023, os quais foram analisados e discutidos. Os resultados preliminares do treinamento e predição demonstraram a eficácia da ferramenta computacional desenvolvida, em relação à assertividade do modelo de predição implementado. Desse modo, pode-se observar a aplicabilidade da Rede Neural Artificial LSTM como uma técnica efetiva para previsão de valores de commodities pecuárias, principalmente da carne bovina.

Palavras-chave – Rede neural artificial, inteligência artificial, predição, agropecuária, análise de dados, séries temporais, rede neural recorrente LSTM.

I. INTRODUÇÃO

Conforme a humanidade evolui é possível visualizar que os processos informatizados estão ficando cada vez mais populares por demonstrarem uma alta versatilidade de capacidade sobre as coisas que se podem fazer com eles, sendo uma delas a Inteligência Artificial.

As redes neurais artificiais (RNA) e a capacidade computacional contemporânea vêm evoluindo juntas, o que

possibilita avanços tecnológicos importantes para diversos processos, industriais, econômicos, urbanos etc. Diante disso, o artigo presente, demonstra um estudo aplicado no que abrange os conceitos, características e princípios relacionados a estas redes.

Junto ao estudo e aplicação de redes neurais artificiais, é necessário entender as séries temporais, que são um instrumento que auxilia e fornece informações que trazem entendimento do mercado para análises e formulações de planejamento, podendo abranger quaisquer setores econômicos. Desta forma, séries temporais são utilizadas para demonstrar padrões que possibilitem previsões que, devidamente estudadas e entendidas, podem trazer alto grau de assertividade.

Um exemplo disso é o artigo da FGV EESP que utiliza um modelo de Redes Neurais Recorrentes LSTM (*Long Short Term Memory*) para prever preços de energia elétrica no mercado brasileiro, considerando fatores históricos e de performance de mercado [1]. Outro estudo, também realizado pela FGV EESP, indica que a técnica RNN LSTM é eficiente para prever vendas no varejo de moda, auxiliando as empresas a entender a sazonalidade do comércio [2]. Além disso, é possível prever a velocidade do vento na região nordeste do Brasil usando séries temporais, permitindo uma instalação mais eficiente de usinas eólicas [3].

Diante a isso, este estudo tem como objetivo analisar séries temporais, utilizando uma RNA do modelo redes neurais recorrentes do tipo LSTM (*Long Short Term Memory*) que é capaz de fazer predições futuras dos preços referentes ao boi gordo, ou seja, treinar a rede neural para entender, analisar e retornar dados de preços futuros no setor agropecuário utilizando informações históricas extraídas de bancos de dados.

Portanto é fundamental destacar a importância das predições que serão realizadas pela rede. Com a capacidade de extrair respostas probabilísticas futuras, traz a possibilidade de planejamento e reorganização de todo o setor envolvido na produção, logística e venda relacionada à commodity do boi gordo.

A aplicabilidade das Redes Neurais Artificiais (RNA) se dá por um sistema voltado para a predição de parâmetros do setor agropecuário utilizando padrões de comportamento da série temporal. As RNA's podem ser vistas como importantíssimas para o processamento das séries, procurando alcançar alta acurácia via relações elementares computacionais. Os modelos das RNA's implicam em exercer de forma contínua, conjecturas utilizando normas e princípios de aprendizagem que aderem poder de discernimento no que tange os valores dos dados da série trabalhados pela rede.

Nas respectivas seções serão apresentadas etapas estudadas para entender a decisão dos modelos computacionais para o processo de predição do preço do boi

gordo. A seção II evidencia as fundamentações teóricas abrangendo o conteúdo necessário para entendimento: do mercado de commodities, das técnicas de inteligência artificial e das séries temporais. A seção III descreve os materiais e métodos utilizados, onde apresenta o banco de dados que será utilizado, bem como as ferramentas computacionais envolvidas no processo. A seção IV abrange os resultados obtidos nas simulações sobre o desenvolvimento e discussões dos dados. As conclusões desse trabalho final de curso 2 podem ser conferidos na seção V.

II. FUNDAMENTAÇÃO TEÓRICA

A Inteligência Artificial é uma ciência que surgiu após a Segunda Guerra Mundial e que tem como principal aplicação a realização de tarefas automatizadas que envolvam certo nível de inteligência, tarefas essas que podem ser simples, como um jogo de xadrez, ou tarefas mais complexas, como diagnósticos de doenças [4].

Sua fundamentação vem a partir de um aglomerado de ideias de diversas outras áreas de conhecimento (filosofia, matemática, economia, neurociência, psicologia, engenharia e computação) contribuindo significativamente para seu desenvolvimento [4].

Atualmente a IA costuma ser aplicada em projetos que possuem como finalidade permitir que máquinas tenham “inteligência” própria suficiente para resolver um problema sem depender de uma pessoa para resolvê-lo. Essa máquina é codificada com algoritmos que tem por finalidade reproduzir a forma como os seres humanos pensam para resolver os problemas propostos a ela.

Devido à grande variação de formas diferentes com que o homem pensa, também é possível afirmar que não seria diferente com a IA, logo criou-se então o que é chamado de “Técnicas de Inteligência Artificial” que são nada mais do que formas diferentes de se aplicar IA em um projeto.

Essas técnicas podem ser ditas em dois grupos: A Inteligência Artificial Clássica, onde estão localizados os sistemas especialistas e os sistemas que são baseados em inferência lógica; e a Inteligência Computacional, ou como é mais popularmente conhecida, o Aprendizado de Máquina (*Machine Learning*) [4].

O sistema inteligente de apoio a decisão (SIAD) desenvolvido nesse trabalho está baseado na técnica de aprendizado de máquina conhecida como redes neurais artificiais recorrentes. A modelagem da base de dados, com as informações das commodities ao longo de um período, é realizada através de séries temporais.

A. *Commodities*

O significado da palavra *commodity* em si se trata de mercadoria, porém, na atualidade, não se trata de qualquer mercadoria, mas sim em produtos básicos de matéria prima. Os produtos que são entregues e produzidos em larga escala e sua precificação é baseada de acordo com a oferta e procura, referente ao setor que está incluído a *commodity* [5].

O Brasil é um dos principais países que se destacam quando se refere a produção de *commodities*, principalmente no que tange aos produtos de origem agrícola e pecuária. Os sete principais produtos produzidos em larga escala pelo território brasileiro são: soja, minério de ferro, petróleo bruto, açúcar, boi gordo, celulose e milho [5].

Tipos de *Commodities*

O ICB (Índice de *Commodities* Brasileiro) pode ser composto de quaisquer *commodities* - produto agrícola, pecuário ou florestal, metais ou energia [6].

Esse trabalho estuda *commodities* relacionadas ao setor agropecuário, mais especificamente no que tange a indicadores de precificação de carne. A produção é o principal fator que altera o preço de quaisquer tipos desse produto, sendo tanto de espécies diferentes (como bovino, suíno, aves, peixes etc.), assim como pode variar de acordo com a raça do animal e tipo do corte.

Levando para a área mais comum no nosso país, que no caso se trata de carne bovina, estudar e compreender indicadores é imprescindível, para o mercado nacional e até mesmo internacional, tendo em vista que o Brasil é um grande exportador de carne. Portanto conseguir apresentar dados sólidos no que se refere à projeção da produção e da precificação dessa *commodity* pode alavancar ainda mais a economia do setor. Isso possibilita melhorias no processo de produção, além de oferecer melhores alternativas de comercialização por parte do produtor de boi gordo.

O estudo das *commodities* exige conhecer dados de produção, precificação de mercado, entre outras. Geralmente, essas informações são registradas ao longo do tempo. Consequentemente, a análise preditiva de qualquer *commodity* necessita, portanto, de modelagem dos dados através de séries temporais, fazendo que o estudo delas demandem uma análise aprofundada das informações disponíveis, possibilitando um melhor entendimento do mercado e tomada de decisões mais assertivas.

B. *Séries Temporais*

Séries temporais são um conjunto de análise de informações, que apresentam dependência consecutiva e serial, ou seja, são informações que não são independentes durante um determinado período. Os ciclos são, de maneira repetida e suavizada, composições de oscilações de queda e alta nas respectivas tendências geradas pelos dados obtidos. A formação de padrões sazonais facilita o processo de previsão correspondente às mudanças, no caso o preço, de oscilações em um determinado período do mês ou ano [7].

De um modo geral, o que se tem de mais importante são os aspectos da análise e modelagem da série temporal - que é justamente descrever os ciclos, verificação de traços mais relevantes obtidos e relacionar com outras séries, respectivamente - e previsão na série temporal - que a partir de resultados, antes encontrados, inicia o processo de previsão [7].

O russo Nikolai Kondratiev apresentou a ideia de que há fases da economia que se repetem, então é assim que surge o interesse na aplicação de séries temporais para a previsão de preços, tanto no boi gordo, quando na própria peça final entregue ao consumidor, pois o fato da ciclicidade da economia capitalista, possibilita uma previsão na oscilação no que diz respeito à precificação dessas *comodities* [8].

Elementos de Séries Temporais

Entre os elementos que compõem as series temporais destacam-se:

- *Tendência*: definida por padrões de oscilações durante a série, variando de acordo com o tempo. A técnica mais

utilizada quando se trata de tendência é a regressão linear simples [9].

- **Sazonalidade:** definida por padrões de comportamento da série no que se refere às épocas do ano. Como por exemplo, preços do litro de gasolina, preço de hospedagem de hotel, preço de peças de carne etc. [9].
- **Estacionariedade:** refere a uma propriedade estatística desejável de uma série temporal, em que a média, a variância e a auto covariância (ou correlação entre observações separadas por uma determinada defasagem) são constantes ao longo do tempo [9].
- **Autocorrelação:** definida por meio de padrões durante um determinado período, relacionando variáveis dos dados em um determinado *gap*, conseguindo correlacionar dados passados para predição futura. Como por exemplo rendimento de um mercado, uma quantia determinada foi vendida no mês de março, é relacionada com a quantia de rendimento de fevereiro, que por sua vez é relacionada a de janeiro [9].
- **Variações Atípicas:** são valores que se afastam significativamente do padrão da série temporal, valores estes que podem ser dados por eventos extremos ou mudanças no comportamento da série [10].

Aplicações de Séries Temporais

Séries temporais tem suas aplicações em diversas áreas, desde identificação de barulhos até previsão de queda e alta da bolsa de valores. Uma sequência de pontos ordenados de forma cronológica possibilita identificação dos dados e principalmente as respectivas predições. Podem ser destacadas três aplicações de séries temporais:

- **Classificação:** aplicação onde há a necessidade de identificação de padrões da série em um determinado instante [11].

- **Agrupamento:** mais comumente reconhecido como “*clustering*”, é uma aplicação que surge da necessidade de agrupar pontos em uma base de dados, analisando a proximidade entre os pontos encontrados no estudo. Desta forma ao identificar os grupos, é possível identificar o que cada um representa [11].

- **Previsão:** quando se trata de valores futuros, há a necessidade de “enxergar” um padrão que permita entender e estudar qual a tendência da série temporal [11]. No caso desse trabalho, destaca-se qual a tendência (predição) na precificação da *commodity* do boi gordo.

C. Técnicas de Predição de Séries Temporais

Atualmente, as séries temporais são amplamente utilizadas em várias aplicações de setores diferentes, se mostrando flexível e fundamental para análise de dados e entregas de respostas. Para chegar nessa eficiência, houve evoluções nas técnicas, unindo redes neurais e séries temporais, que facilitam a predição, principalmente de *commodities*, para análise de informações e, a partir disso, possibilitando assertividade nas previsões de eventos futuros nos setores.

Os modelos de predição se baseiam numa função já ocorrida no passado, conseguindo projeções para o futuro, ou seja, capturando informações sobre o que já ocorreu por meio de análise de eventos acontecidos e utilizando uma série de dados para fazer uma previsão. Entre as técnicas de predição

de séries temporais, podem ser destacadas três: o ARIMA, o SVM e a RNA.

ARIMA

A técnica ARIMA (*Auto Regressive Integrated Moving Average*) é constituída basicamente pela utilização de dados passados para a previsão futura, utilizando a autocorrelação e médias móveis como recursos para o processo de predição [12].

A técnica ARIMA possui três parâmetros principais, que são P, D e Q. Cada letra representa uma parte da sigla ARIMA conforme especificado na Figura 1.

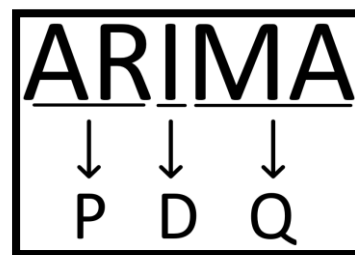


Figura 1 – Entendendo P, D e Q da técnica ARIMA [8].

Os termos P, D e Q são notações que representam respectivamente, número de autorregressão, número de diferenciação para a estacionariedade e, por último, o número de parâmetro de médias móveis [7]. Portanto, ao construir o modelo ARIMA, é necessário estabelecer os parâmetros em forma de tupla, no qual, “p”, “d” e “q” são números inteiros, como por exemplo (1,1,1), (0,1,0), (2,0,1) e assim por diante [12].

A utilização do ARIMA exige o conhecimento de duas funções imprescindíveis, que são, a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF). Em uma análise de série temporal, a correlação dá informações de valores de um dado com outro valor, a autocorrelação é a comparação de um valor do presente com um valor do passado, na mesma série temporal. Já a autocorreção parcial (PACF), se diferencia fazendo apenas correlação direta, enquanto a autocorrelação (ACF), faz relações diretas e indiretas. Exemplificando em relação ao preço da *commodity* boi gordo: utilizando a (ACF) consegue-se a correlação direta dos preços do mês de agosto em outubro e a correlação indireta que o mês de agosto teve em setembro e que também teve em outubro [12].

Máquina de vetores de suporte (SVM)

Uma máquina de vetores de suporte (SVM, do inglês *support-vector machine*) é um conjunto de métodos de aprendizagem supervisionado. Esse método reconhece e analisa padrões por análise de regressão em que o SVM padrão tem como entrada um conjunto de dados, que é extraído de um repositório de registros, e dessa forma consegue prever, para todas as entradas, quais as possíveis classes que as entradas podem fazer parte. Ou seja, o SVM pode ser visto como um classificador linear binário não probabilístico [13].

O SVM tem como característica, um ótimo funcionamento com separação de dados quando se trata de valores bastante distintos, ou seja, na discrepância de valores de dados há uma eficiência que facilita a análise. Porém quando se trata de uma grande base de dados, em que o conjunto de amostra é muito grande, essa técnica é

inadequada, pois esse grande conjunto de dados demandaria um tempo muito grande de treinamento [13].

Ao utilizar SVM para a regressão, é possível analisar valores de acordo com o padrão da série, e desta forma conseguindo previsões dos futuros valores da série [14].

A Figura 2 descreve, de forma simplificada, como a técnica SVM trabalha. Ao considerar um conjunto de dados (quadrados em vermelho), e ao usar regressão, considera-se os pontos mais distantes (dados com maiores discrepâncias de valores), respeitando uma tolerância ϵ (épsilon), distância da linha tracejada para a linha central do gráfico. Então, a tarefa atribuída ao SVM é encontrar a reta central, que é a reta mais próxima desses dados externos que respeitam ϵ [13].

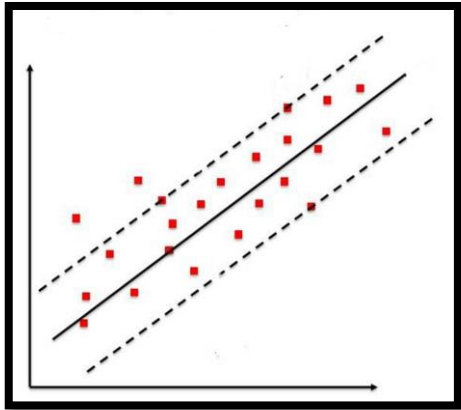


Figura 2 – SVM – Modelo de previsão simplificado [10].

Redes Neurais Artificiais (RNA)

A inteligência do homem é a mais desenvolvida dentre todo o universo até então conhecido. O cérebro é quem possui as devidas ferramentas para obter tamanho desenvolvimento e diferenciação dentre outras espécies, no caso as ferramentas mais básicas que possuímos são os neurônios, que se conectam em redes que permitem a circulação de informações entre eles. O homem criou essa complexa rede que possuímos, de uma forma artificial, pelo fato de ser a melhor já apresentada na natureza, criando a estrutura e funcionamento do cérebro, só que dessa vez em um ambiente técnico [14].

A estrutura de trabalho de uma rede neural artificial se dá por séries de processos, formando dessa forma, uma rede de dados [14]. Para o funcionamento das redes neurais artificiais, é necessário que haja três camadas integradoras na rede, sendo elas:

- Camada de Entrada: Camada cuja os valores dos dados da série temporal são fornecidos para a RNA;
- Camada Oculta: Camada responsável por todo o processamento, do qual identifica as características da série avaliada;
- Camada de Saída: Camada em que os dados são retornados para apresentação do resultado da previsão.

Seguindo o respectivo cronograma da RNA (camada de entrada, camada oculta e em seguida camada de saída), é basicamente todo o processo que deverá ser respeitado para o funcionamento da rede [1].

Além do que se refere ao funcionamento de uma RNA, outra etapa indispensável, é a habilidade de treinamento, que é quando pode-se aprender dados, características e informações, por meio de memorização, analogia, regressão e análise. Desta forma há dois tipos de aprendizagem, que é a

maneira em que a RNA se relaciona diante aos eventos do ambiente, sendo:

- Aprendizado Supervisionado: modelo em que é usado um padrão em que já se sabe qual o resultado deverá ser retornado como saída da rede;
- Aprendizado Não Supervisionado: modelo em que é usado um padrão, porém desta vez, não há se sabe o resultado que deverá ser apresentado [1].

D. Rede Neural Recorrente LSTM

A LSTM (*Long Short Term Memory*) é uma arquitetura de rede neural bastante adequada para classificações, processos e previsões temporais com períodos de duração desconhecida. A LSTM por desconsiderar o tamanho do intervalo fica à frente no que se refere à confiabilidade em relação às RNNs (*Recurrent Neural Networks*) tradicionais [15, p. 31].

A LSTM constitui de uma cadeia de dados que contém quatro redes neurais diferentes e distintos blocos de memórias, bem como mostrado na Figura 3, onde:

Forget Gate: Os dados que não são mais necessários são apagados. O resultante se dá por uma saída binária, no qual é determinado o estado de célula, que em caso de saída 0 é removida e em caso de saída 1 o dado permanece no processo, sendo utilizado futuramente [15];

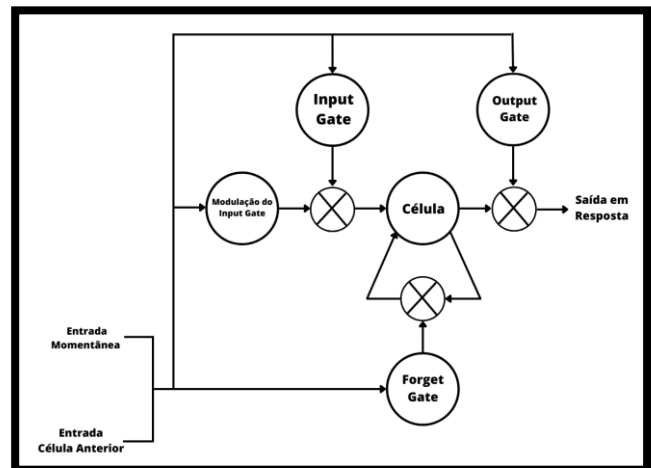


Figura 3 – Imagem explicativa das células referentes à LSTM [1].

Input Gate: Para a entrada de dados úteis a informação é filtrada para valores a serem lembrados, utilizando sigmoide (função real, diferencial e limitada, $D(f) = R$, cuja valor entregue está incluso ao intervalo (0,1) possuindo derivada positiva em cada ponto). Dessa forma contendo todos os valores possíveis, sendo multiplicado pelos valores regulado, para assim então, encontrar informações uteis [15].

Output Gate: A função é de extrair informações úteis da célula atual para serem apresentadas como saída. O vetor é aplicado, e então a informação é regulada, para que seja guardada. Os valores da célula são multiplicados para serem enviados como saída e entrada da célula seguinte [15].

A construção de uma RNN LSTM, é, normalmente, dividida em duas bases de dados de entrada, que são capazes de treinar a respectiva IA. As propriedades são as mesmas das

redes neurais convencionais, porém com um fator somatório que consegue armazenar informações por grandes períodos ao computar dados em sequência temporal, por possuir células, que são capazes de “lembrar” as informações necessárias e úteis [15].

A informação de entrada na célula é uma decisão obtida pelo *Input Gate*. O *Forget Gate* define a utilidade do dado, sendo assim, é ele quem define o descarte da informação. O estado de processamento é memorizado, e todas as informações úteis - haja visto que as não úteis foram descartadas no *Forget Gate* – serão computados pelo *Output Gate* [15].

III. MATERIAIS E MÉTODOS

Com o intuito de projetar um sistema que possa resolver problemas sem a necessidade da intervenção humana e com uma performance igual ou até mesmo superior a ela, foi realizado um estudo para criar uma inteligência artificial (IA) utilizando uma rede neural recorrente LSTM. Esse estudo foi dividido em duas partes: a busca da base de dados que é utilizada no trabalho e a montagem do sistema de apoio a decisão com RNN LSTM, que utiliza essa base de dados para realizar o processo de predição. Essa rede neural é responsável por realizar a predição dos valores da carne bovina para que seja possível realizar a verificação e uma estimativa de preço que ela terá no futuro. Serão utilizados os valores históricos e atuais dessa *commodity* como forma de treinamento para que a rede seja capaz de realizar as predições futuras.

A. Base de Dados

A base de dados utilizada nesse trabalho é composta por uma série temporal com os valores dos indicadores do Boi Gordo CEPEA/B3, em reais e em dólares, que foram obtidos junto ao Centro de Estudos Avançados em Economia Aplicada (CEPEA), em parceria com a Escola Superior de Agricultura Luiz de Queiroz (ESALQ) da Universidade de São Paulo (USP). Os valores utilizados foram obtidos diariamente, com exceção dos finais de semana, do dia 23 de julho de 1997 até o dia 26 de maio de 2023, com um total de 6427 amostras [16].

Etapa de Inicialização

A etapa de inicialização é realizada utilizando o arquivo *Microsoft Excel* contendo a base de dados, que pode ser carregada em um *notebook* do *Jupyter Notebooks* utilizando a IDE *Visual Studio Code (VS Code)*. Esse procedimento possibilita que a base de dados seja utilizada conforme desejado na aplicação em questão. Esse carregamento é realizado utilizando a biblioteca *Pandas*, que é uma biblioteca responsável por realizar o *upload* e trabalhar com arquivos de base de dados (podendo esses serem “.xlsx” ou “.csv”).

Padronização dos Dados

Para que os dados possam ser utilizados e processados corretamente em uma plataforma de desenvolvimento, muitas vezes é necessário realizar o processo de padronização. Isso evita possíveis distorções durante a etapa de treinamento da rede neural LSTM, garantindo que os dados estejam em um formato adequado para serem utilizados na análise. Neste trabalho, os dados coletados do site do CEPEA/ESALQ-USP foram transformados em um vetor e padronizados com o

auxílio da biblioteca *Scikit Learn* e do comando *'StandardScaler()'*, que foram utilizados juntamente com a função *'scaler.fit_transform()'*. A padronização foi realizada por meio de um cálculo que resultou em uma pontuação para cada dado, conforme a Equação (1):

$$z = (x - u)s^{-1}, \quad (1)$$

Onde: u representa a média das amostras de treinamento (ou 0 caso não exista) e s é o desvio padrão dessa amostra (ou 1 caso não exista) e x é o valor que está sendo lido pela função. Os valores são armazenados e utilizados durante todo o processo, até depois ser feita a despadronização dos dados com a função *"scaler.inverse_transform()"*[17].

B. RNN LSTM – Implementação Computacional

A linguagem de programação Python vem se tornando cada vez mais conhecida por ser uma linguagem extremamente dinâmica. Desde o começo dos anos 2000 ela está se tornando cada vez mais ativa e utilizada para ciência de dados tanto na indústria, quanto em aplicações acadêmicas [18].

O fator que levou a escolha da utilização dessa linguagem é a facilidade de aprendizagem da linguagem e a grande variedade de bibliotecas que ela possui e que são voltadas para análise de dados, tornando-a a linguagem ideal para a realização de análise de séries temporais.

Configuração da RNN LSTM

Nesse trabalho será utilizado como configuração de redes neurais um modelo de LSTM (*Long-Short Term Memory*) a função de perda calculada pela raiz do erro quadrático médio (*Root Mean Square Root Error* ou *RMSE*) e o otimizador *'Adam'* (*Adaptive Moment Estimation*).

Treinamento da RNN LSTM

O treinamento da LSTM é feito em duas etapas. A primeira etapa treina a sequência de dados de julho de 1997 até julho de 2019 apresentando o melhor resultado com a taxa de aprendizado de 0,001, 300 épocas, um lote de 32 amostras e utilizando 20 neurônios. A segunda etapa treina a sequência de dados de julho de 2019 até maio de 2023 e apresenta seu melhor resultado com a taxa de aprendizado de 0,001, 500 épocas, um lote de 16 amostras e utilizando 15 neurônios. As Tabelas 1 e 2 mostram alguns resultados de treinamentos realizados com a rede afim de obter a melhor combinação dos hiper parâmetros para cada uma das etapas.

Tabela 1 – Tabela de valores comparando erros com variações de hiper parâmetros – 1997/2019.

Tabela de valores comparando erros com variações de hiperparâmetros - 1997/2019							
	Épocas	Taxa	Lote	Neurônios	R ²	RMSE TRAIN	RMSE TEST
1	100	0,5	10	5	-4,79%	5,34	50,02
2	100	0,01	16	5	60,00%	1,08	13,19
3	100	0,001	16	10	97,00%	0,38	3,42
4	200	0,001	16	10	98,00%	0,38	3,19
5	200	0,001	16	15	98,00%	0,35	3,11
6	200	0,00001	16	15	91,00%	1,66	6,15
7	200	0,0001	16	15	89,00%	0,57	6,99
8	200	0,001	32	15	99,00%	0,47	2,51
9	200	0,001	32	20	99,00%	0,4	2,3
10	300	0,001	32	20	99,00%	0,35	1,93
11	300	0,00001	16	15	83,00%	2,07	8,69

Tabela 2 – Tabela de valores comparando erros com variações de hiperparâmetros – 2019/2023

Tabela de valores comparando erros com variações de hiperparâmetros - 2019/2023							
	Épocas	Taxa	Lote	Neurônios	R ²	RMSE TRAIN	RMSE TEST
1	100	0,5	10	5	35,00%	13,42	18,09
2	100	0,01	16	5	91,00%	4,07	6,88
3	100	0,001	16	10	90,00%	4,94	7,17
4	200	0,001	16	10	91,00%	3,77	6,9
5	200	0,001	16	15	90,00%	3,75	6,93
6	300	0,001	16	15	91,00%	3,6	6,74
7	500	0,001	16	15	91,00%	3,59	6,71
8	350	0,001	32	20	91,00%	3,72	6,9
9	500	0,001	32	20	91,00%	3,66	6,85
10	500	0,0001	32	20	90,00%	5,17	7,07
11	800	0,0001	32	20	90,00%	3,79	6,96
12	1000	0,0001	32	20	90,00%	3,82	6,96
13	1000	0,0001	32	40	90,00%	3,76	6,91

Nas duas tabelas, as linhas selecionadas com a cor verde indicam a configuração com o melhor resultado para a rede considerando as métricas R² e o RMSE. Em ambos os cenários se utiliza 67% da base de dados para treinamento e 33% para validação.

C. Cenário de Aplicação

O objetivo do sistema de predição é identificar os valores futuros de mercado para o custo da arroba do boi gordo. Para alcançar esse objetivo, primeiramente é de suma importância conhecer bem o histórico de registros da base de dados utilizada. Portanto, nesse trabalho, os estudos procuram analisar quantitativamente o comportamento das séries temporais que descrevem a variável em questão.

Nesse trabalho final de curso 2 o cenário de aplicação representa o estudo da série temporal propriamente dita, com apontamentos de tendência, sazonalidade e comportamentos atípicos. O objetivo é levantar subsídios para melhor ajustar os parâmetros da rede LSTM utilizada no processo de predição.

Alguns cenários de predição também são avaliados nesse trabalho, embora as análises ainda sejam bastante superficiais e careçam de estudos mais aprofundados.

A Figura 4 descreve o cenário de aplicação da rede LSTM. O Sistema de predição pode receber informações de registros, e se for treinado adequadamente, é capaz de prever o custo futuro da arroba do boi gordo, para a semana seguinte, por exemplo.

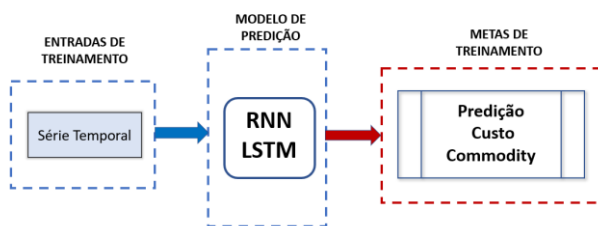


Figura 4 – Treinamento do sistema de predição de custo de commodity.

VI. RESULTADOS

Nesta seção serão descritos os resultados obtidos ao longo deste trabalho final de curso 2 com a finalidade de pontuar os valores encontrados nas seções de configurações e treinamento da RNN LSTM e analisar a viabilidade da aplicação dessa rede neural para a predição da precificação do valor da carne bovina, esperando uma resposta positiva sobre essa aplicação.

A Figura 5 apresenta o resultado do carregamento da base de dados na plataforma de desenvolvimento utilizada.

Base de Dados - Boi Gordo - CEPEA			
	Data	Preço R\$	Preço US\$
0	23/07/1997	26,67	24,65
1	24/07/1997	26,67	24,65
2	25/07/1997	26,71	24,68
3	28/07/1997	26,74	24,70
4	29/07/1997	26,77	24,72
...
6422	22/05/2023	263,55	53,04
6423	23/05/2023	261,85	52,66
6424	24/05/2023	257,30	51,96
6425	25/05/2023	255,85	50,78
6426	26/05/2023	253,30	50,76

Figura 5 – Imagem da base de dados carregada no Jupyter Notebook utilizando o Visual Studio Code.

Nesses dados pode-se verificar a grande diferença que o preço, em reais, da arroba do boi gordo sofreu, tendo um aumento de quase de 950% do seu valor inicial. Esse crescimento pode ser explicado utilizando fatores econômicos como inflação, depreciação da moeda brasileira e emergências ocorridas ao longo dos anos que geram o aumento da procura dessa commodity, aumentando também o seu valor econômico.

Com o objetivo de verificar melhor o comportamento da série temporal da commodity, foi gerado um gráfico que descreve o custo da arroba do boi gordo ao longo do período de estudo. A Figura 6 mostra esse gráfico, descrevendo a série temporal para o custo da commodity em reais.

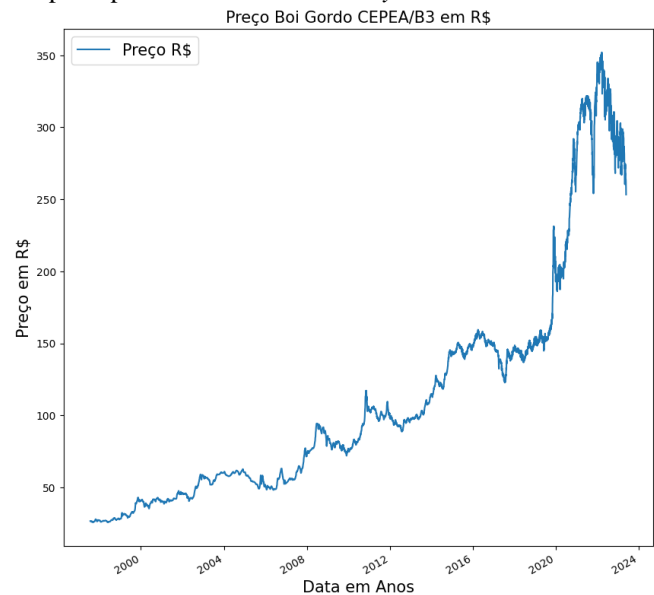


Figura 6 – Valores do indicador da arroba do Boi Gordo CEPEA/B3 de julho/1997 até maio/2023 em reais.

No gráfico apresentado na Figura 6 é possível identificar um pico próximo ao final do ano de 2019 e início de 2020, fato esse que se dá devido ao aumento de preço que se teve em virtude da pandemia do vírus da COVID-19. Esse comportamento totalmente atípico da série temporal merece um estudo aprofundado, com destaque para o grau de influência desse evento no sistema de predição desenvolvido. Por conta dessa alta variação, a série temporal foi dividida em dois períodos: um de 23 julho de 1997 até 23 de julho de 2019 e o outro de 24 de julho de 2019 até 26 de maio de 2023. As Figuras 7 e 8 descrevem esse detalhamento dos dados.

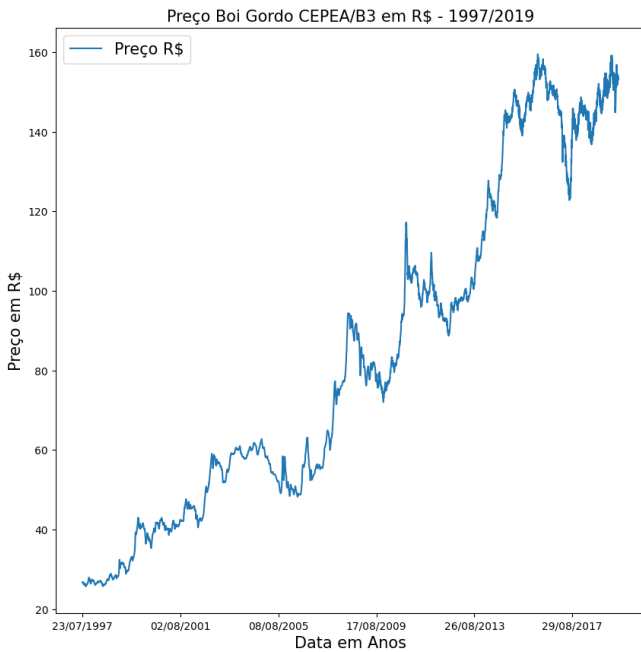


Figura 7 – Valores do indicador da arroba do Boi Gordo CEPEA/B3 de julho/1997 até julho/2019 em reais.

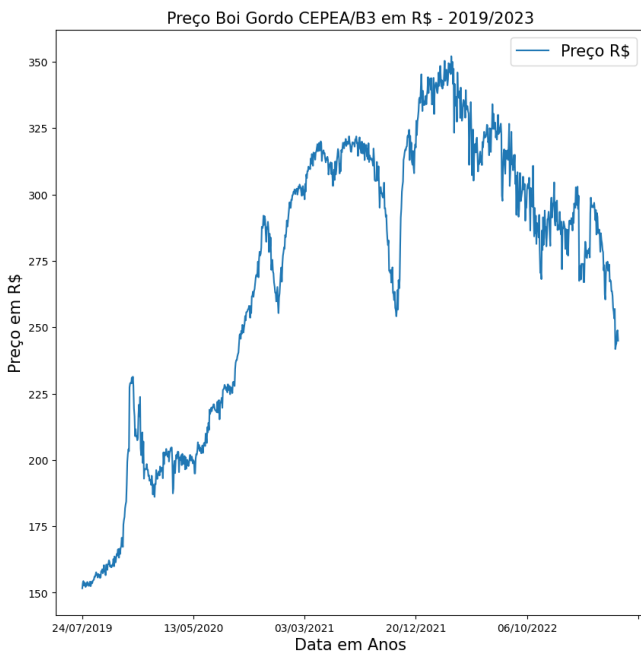


Figura 8 – Valores do indicador da arroba do Boi Gordo CEPEA/B3 de julho/2019 até maio/2023 em reais.

Como é possível observar após a divisão dos valores utilizados na base de dados, ambos os gráficos, apresentados nas Figuras 7 e 8, possuem uma forma que facilita a leitura e predição dos valores expostos por conta da linearidade evidenciada, além de separar os dados de forma que os valores obtidos durante a pandemia da COVID-19 (Figura 8) fiquem isolados, facilitando o estudo que pode ser realizado sobre a influência desse evento na precificação da arroba do boi gordo.

Com o conhecimento do comportamento da série temporal da commodity, foram realizadas simulações computacionais para verificar a eficácia da técnica RNN LSTM empregada neste trabalho no processo de predição. O objetivo é verificar a eficácia da ferramenta computacional implementada na predição do custo da arroba do boi gordo para os períodos de análise da série.

A rede neural LSTM foi testada em diversos cenários de configuração, com ajustes em alguns de seus parâmetros, com destaque para o número de épocas, o tamanho do lote e a taxa de aprendizado e quantidade de neurônios sendo utilizados, como foram apresentados na seção de “*Treinamento da RNN LSTM*”. Os melhores resultados alcançados são apresentados nas Figuras 9 e 10, onde o eixo horizontal representa a data da coleta da amostra em questão enquanto que o eixo vertical representa o valor da arroba, em reais; A curva azul representa os dados originais da base de dados (Figura 5), a curva laranja os valores da predição realizados durante o treinamento da LSTM e a curva verde os valores que foram validados pela LSTM, ou seja, os valores de teste previstos. Nesses cenários, a LSTM foi configurada para a predição do preço do dia seguinte a partir da leitura dos preços de 10 dias anteriores.

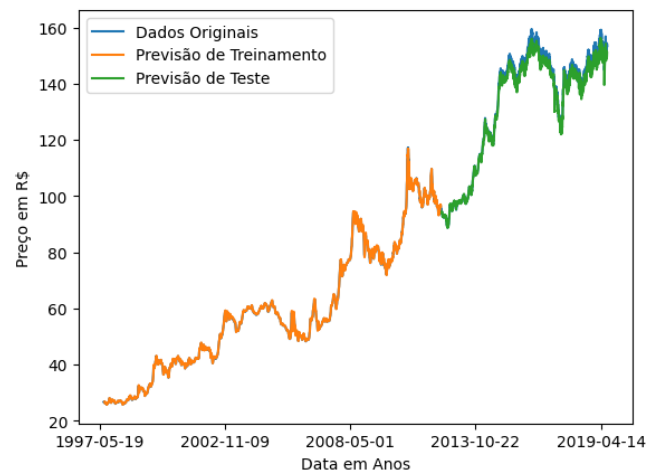


Figura 9 – Resultado da predição dos valores da arroba do boi gordo utilizando os valores de julho/1997 até julho/2019.

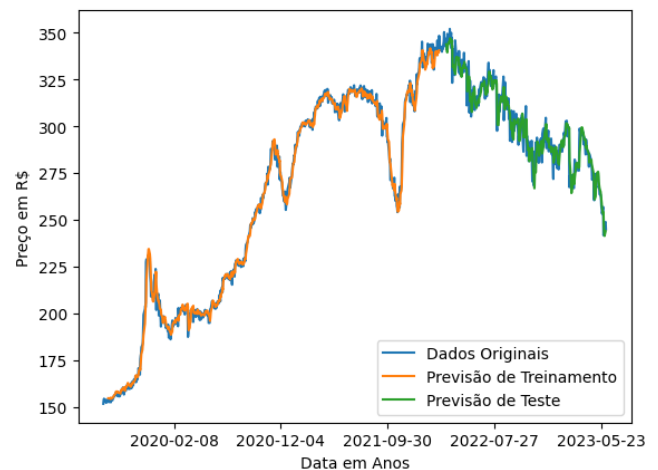


Figura 10 – Resultado da predição dos valores da arroba do boi gordo utilizando os valores de julho/2019 até maio/2023.

O cenário pré-pandemia da COVID-19 (Figura 9) apresentou uma acurácia de 99%, e RMSE de 0,35 na etapa de treinamento e 1,93 na etapa de validação. Esses resultados evidenciam uma alta precisão do processo preditivo. Já, durante o período da pandemia (Figura 10), a LSTM apresenta maior dificuldade na predição, com uma acurácia de 91% e um RMSE de 3,59 nos dados de treinamento e 6,71 nos dados de validação. Essa dificuldade de predição da LSTM em capturar a tendência da série temporal no cenário de variações atípicas deverá ser amenizada à medida que se

tenha uma quantidade maior de registros pós pandemia que possam ser agregados ao processo de treinamento.

Apesar dessas limitações, é importante ressaltar que ambos os resultados de predição são consistentes com as expectativas e ainda são considerados muito bons. Mesmo durante a pandemia, em que a assertividade é menor, a precisão da previsão ainda é significativa.

VII. CONCLUSÃO

Conforme a sociedade evolui, novas tecnologias vão surgindo para acompanhar esse desenvolvimento, sendo uma delas a Inteligência Artificial, que teve como um de seus objetivos produzir diversos SIAD (Sistemas Inteligentes de Apoio a Decisão) que pudessem ser utilizados como alternativas para a automação de processos e tomada de decisões.

Nesse viés, é imprescindível a utilização de técnicas de IA para facilitação de leitura e análise de dados, visto que a quantidade de dados disponíveis atualmente pode ser exorbitante a depender da aplicação. Esse trabalho teve por finalidade realizar um estudo inicial sobre a aplicação de uma dessas técnicas, a rede neural recorrente com memória a longo e a curto prazo (RNN LSTM) para a leitura, análise e predição de dados.

Para os cenários avaliados nesse trabalho, os resultados obtidos no processo de predição do valor da arroba do boi gordo, se mostraram bastante promissores. Embora a acurácia da predição seja considerada boa, ainda se torna necessário realizar simulações computacionais com outros modelos de predição e aprimorar o processo de ajustes nos parâmetros de configuração da rede LSTM.

Portanto, a aplicação de técnicas de IA, como a RNN LSTM, pode trazer grandes benefícios para a predição de commodities, tornando possível a automação de processos e a tomada de decisões com maior precisão. Nesse sentido, é importante continuar os estudos sobre essa aplicação, aprimorando-a e buscando soluções para os desafios encontrados, a fim de obter resultados cada vez mais satisfatórios e precisos.

VIII. REFERÊNCIAS

- [1] Santos, G.: Uma aplicação de redes neurais recorrentes do tipo LSTM à previsão dos preços de curto prazo do mercado de energia elétrica brasileiro. Dissertação de mestrado. São Paulo: Fundação Getúlio Vargas (2019).
- [2] Bezerra, A.: Previsão de vendas no varejo de moda com modelos de redes neurais. Dissertação de mestrado. São Paulo: Fundação Getúlio Vargas (2018).
- [3] Camelo, H.d.N., Lucio, P.S., Leal Junior, J.B.V., Carvalho, P.C.M.d.: Métodos de previsão de séries temporais e modelagem híbrida ambos aplicados em médias mensais de velocidade do vento para regiões do nordeste do Brasil. *Revista Brasileira de Meteorologia* 32(4), 565-574 (2017).
- [4] Russell, S., & Norvig, P. (2004). *Inteligência Artificial* (2ª ed.). Rio de Janeiro: Elsevier.
- [5] Abreu, J. (s.d.). O que são commodities e quais são os tipos? Serasa. Recuperado em 25 de novembro de 2022, de <https://www.serasa.com.br/blog/commodities/>
- [6] B3. (s.d.). Índice de commodities Brasil (ICB B3). Recuperado em 13 de novembro de 2022, de https://www.b3.com.br/pt_br/
- [7] Veríssimo, J. L.: *Séries Temporais: Um estudo de Previsão* (Trabalho Final de Curso). Universidade Federal da Grande Dourados (2016).
- [8] Kondratieff, N.: *The Long Waves in Economic Life. The Review of Economic Statistics*, The MIT Press, v. 17, n. 6, p. 105-115, nov., 1935.
- [9] Latorre, M.d.R.D.d.O; Cardoso, M.R.A.: Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*, São Paulo, v. 4, n. 3, p. 145-152, 2001.
- [10] Stevenson, W.J.: *Estatística Aplicada a Administração*. São Paulo: Harbra, 1981.
- [11] Hastie, T.; Tibshirani, R.; Friedman, J.: *Os Elementos do Aprendizado Estatístico: Mineração de Dados, Inferência e Predição*. 2ª ed. São Paulo: Bookman, 2013.
- [12] P. Vasconcellos. "Dicas para criar um modelo de previsão de séries temporais". Medium. <https://medium.com/techbloghotmart/dicas-para-criar-um-modelo-de-previsão-de-séries-temporais-d4bb2e32e148> (consult. 2022-11-25).
- [13] D. Addan. "Support vector machine". <https://www.inf.ufpr.br/dagoncalves/IA07.pdf> (consult. 2022-11-28).
- [14] Hochreiter, S.; Schmidhuber, J.: Long Short-Term Memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.
- [15] Data Science Academy. 2022. *Deep Learning Book*. [Versão eletrônica]. Recuperado em 25 de novembro de 2022 de <https://www.deeplearningbook.com.br/>
- [16] Centro de Estudos Avançados em Economia Aplicada - CEPEA-ESALQ/USP. <https://www.cepea.esalq.usp.br/br/indicador/boi-gordo.aspx> (consult. 2022-11-08).
- [17] SCIKIT-LEARN. StandardScaler. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. Acesso em: 05 mai. 2023.
- [18] McKinney, W.: *Python for Data Analysis*. O'Reilly Media, San Francisco (2013).
- [19] G. Trielli Avila, "Regressão em séries temporais financeiras com rnn: Um estudo com milho futuro", Trabalho de Conclusão de Curso, Universidade Estadual Paulista "Júlio de Mesquita Filho", Bauru, 2019.
- [20] H. Eugênio Gonçalves, A. Coutinho Mateus e L. Xavier Medeiros, "Previsão dos índices de chuva acumulados na cidade de Uberlândia utilizando redes neurais artificiais LSTM", in *CEEL*, Uberlândia, Brasil, 2021-12-13–17. 2021.
- [21] A. de Paulo Andrade Filho, "Emprego de redes neurais LSTM na predição de séries temporais aplicadas ao consumo de energia elétrica", in *CEEL*, Uberlândia, Brasil, 2021-12-13–17. 2021.
- [22] K. Ewan Freire Schuch, "Análise preditiva com redes neurais artificiais para o planejamento de sistemas de irrigação", Trabalho Final de Curso, Pontifícia Universidade Católica de Goiás, Goiânia, 2021



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
GABINETE DO REITOR
Av. Universidade, 1089 - Setor Universitário
Cidade de Goiás - GO - CEP 74605-010
Goiânia - Goiás - Brasil
Fone: (62) 3645-1000
www.pucgoias.edu.br - reitoria@pucgoias.edu.br

RESOLUÇÃO nº 038/2020 – CEPE

ANEXO I

APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante FELIPE AUGUSTO ROMANO do Curso de ENGENHARIA DE CONTROLE E AUTOMAÇÃO, matrícula 2019.1.0118.0003-7, telefone: (62)999006200 e-mail feliperomano2846@gmail.com, na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado ANÁLISE PREDITIVA DA PRECIFICAÇÃO DE COMODITIES PECUÁRIAS UTILIZANDO REDES NEURAS LSTM, gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros, específicos da área; para fins de leitura e/ou impressão pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 06 de março de 2023.

Assinatura do autor: *Felipe A. Romano*

Nome completo do autor: Felipe Augusto Romano

Assinatura do professor-orientador: *Marcos A. de Sousa*

Nome completo do professor-orientador: Marcos Antônio de Sousa



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
GABINETE DO REITOR
Av. Universidade, 1089 - Setor Universitário
Cidade de Goiás - GO - CEP 74605-010
Goiânia - Goiás - Brasil
Fone: (62) 3645-1000
www.pucgoias.edu.br - reitoria@pucgoias.edu.br

RESOLUÇÃO nº 038/2020 – CEPE

ANEXO I

APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante MATHEUS FERNANDES RODRIGUES DE SOUSA do Curso de ENGENHARIA DE CONTROLE E AUTOMAÇÃO, matrícula 2019.1.0118.0027-4, telefone: (62) 9 9447-6770 e-mail matheusfernandes1898@gmail.com, na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado ANÁLISE PREDITIVA DA PRECIFICAÇÃO DE COMODITIES PECUÁRIAS UTILIZANDO REDES NEURAS LSTM, gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros, específicos da área; para fins de leitura e/ou impressão pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 06 de março de 2023.

Assinatura do autor: *Matheus Fernandes R. de Sousa*

Nome completo do autor: Matheus Fernandes Rodrigues de Sousa

Assinatura do professor-orientador: *Marcos A. de Sousa*

Nome completo do professor-orientador: Marcos Antônio de Sousa