

UMA ABORDAGEM PARA A PREDIÇÃO DE PREÇOS FUTUROS DE COMMODITIES UTILIZANDO ANÁLISE TÉCNICA E APRENDIZADO DE MÁQUINA

Gabriel Schneider de Jesus, Aurélio Faustino Hoppe – Orientador

Curso de Bacharel em Ciência da Computação
Departamento de Sistemas e Computação
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brasil

gsjesus@furb.br, aureliofof@furb.br

Resumo: Este trabalho apresenta o desenvolvimento e implementação de um modelo de aprendizado de máquina para estimar o preço futuro de commodities no âmbito do mercado brasileiro a partir de indicadores de análise técnica. Para isso, foram obtidas duas bases de dados referentes aos commodities açúcar, algodão, milho, soja e trigo, as quais foram submetidas às etapas de limpeza, pré-processamento e subdivisão de dados. A partir dos dados pré-processados, utilizou-se redes neurais recorrentes do tipo Long Short Term Memory (LSTM) para realizar a previsão de dados no intervalo de 1 e 3 dias à frente. Tais modelos foram avaliados através do Mean Squared Error (MSE), obtendo uma precisão entre 0,00009 e 0,00037 nos dados de teste para 1 dia à frente e de 0,00015 a 0,00041 para 3 dias à frente. Contudo, com base nos resultados obtidos, pode-se afirmar que o modelo desenvolvido obteve uma boa performance de previsão para todas as commodities avaliadas.

Palavras-chave: Commodities. Análise técnica. Previsão. Aprendizado de máquina. Redes neurais. LSTM.

1 INTRODUÇÃO

Commodities são mercadorias produzidas em larga escala, sendo uma de suas principais características o fato de serem estocadas sem perder a qualidade, por exemplo, o petróleo, o café, a soja e o ouro. Tais mercadorias podem ser negociadas no mercado futuro, através de contratos futuros, que representam acordos de compra e venda por um determinado preço e período, semelhante ao mercado de ações (DEMPSTER; MADAN; CONT, 2016; EQUIPE TORO RADAR, 2016).

Segundo Dempster, Madan e Cont (2016), existem duas classes de *traders*, negociadores nos mercados futuros de commodities: *hedgers* e *speculators*. Os *hedgers*, que normalmente são produtores, procuram proteger sua posição física futura em uma mercadoria vendendo contratos futuros. Já os *speculators*, buscam lucros especulativos, assumem o lado da compra desses negócios. Por consequência, realizar previsões precisas sobre o preço de commodities pode beneficiar as duas classes de negociadores, alavancando seu potencial econômico, o que por sua vez pode trazer indiretamente, benefícios a população em geral.

Para Kim (2019), é comumente aceito que existe uma relação inversa entre preços de commodities como petróleo, trigo, metais básicos etc. e a economia: quando os preços das commodities caem, os efeitos econômicos são positivos. Fang *et al.* (2020) ressalta que a previsão de preços de commodities agrícolas é uma tarefa essencial para que o cenário político-econômico de uma nação emergja.

Melhorar a precisão da previsão dos preços futuros dos produtos agrícolas é importante para os investidores, produtores agrícolas e formuladores de políticas. Isso evita riscos e permite aos departamentos governamentais formular regulamentos e políticas agrícolas apropriadas (FANG *et al.*, 2020, p. 1, tradução nossa).

Atualmente, existem diversas pesquisas científicas que investigam a previsibilidade do mercado de commodities bem como do mercado acionário. Dentre elas, pode-se destacar os estudos de Rigatos *et al.* (2018) que propuseram um método para prever preços de commodities usando a Equação Diferencial Parcial (EDP) e filtragem de Kalman. Já Borovkova e Tsiamas (2019) optaram por uma rede neural LSTM para previsões de ações. Por outro lado, na pesquisa feita por Yin e Yang (2016), eles investigaram a usabilidade de indicadores técnicos para prever diretamente os preços do petróleo, comparando seu desempenho com variáveis macroeconômicas. Yin e Yang (2016) ressaltam que os indicadores técnicos possuem previsão superiores à de variáveis macroeconômicas, sendo substancial durante recessões e expansões, podendo detectar de forma eficaz os declínios típicos nos retornos do petróleo perto dos picos dos ciclos de negócios.

Diante destas circunstâncias, este trabalho apresenta o desenvolvimento e implementação de um modelo de aprendizado de máquina que utilize séries históricas de preços internacionais e nacionais, bem como seus respectivos indicadores de análise técnica para estimar o preço futuro de commodities no âmbito do mercado brasileiro, passando pelas etapas de análise dos indicadores técnicos mais suscetíveis ao problema, desenvolvimento de um modelo de previsão

de dados para o mercado internacional e outro para o mercado nacional de commodities e, avaliação de eficiência, gerando por fim, um artefato computacional. Contudo, este trabalho mostra-se relevante pois ajudará a comunidade científica ao apresentar um estudo utilizando algoritmos de aprendizado de máquina em combinação com a análise técnica para a predição de séries temporais à curto prazo, utilizando-se de dados de mercados estrangeiros para a predição de dados nacionais. Por fim, destaca-se sua relevância para investidores, fazendeiros e membros de departamentos governamentais, visto que, com uma predição precisa do preço de commodities é possível tomar boas decisões de compra e venda, podendo assim, alavancar o potencial econômico do indivíduo, empresa ou país.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os principais conceitos e técnicas aplicadas durante o desenvolvimento do trabalho. A seção 2.1 discorre sobre mercado de commodities, análise técnica e indicadores técnicos. A seção 2.2 por sua vez, conceitua aprendizado de máquina e as redes neurais com arquitetura *LSTM*.

2.1 MERCADO DE COMMODITIES, ANÁLISE TÉCNICA E INDICADORES TÉCNICOS

As *commodities* sempre fizeram parte do nosso dia a dia e, agora, também são consideradas uma importante classe de ativos nos mercados financeiros. A maioria das *commodities*, incluindo o ouro, o petróleo, o gás natural etc., agora são negociadas por contratos futuros em bolsas de valores para produtos agrícolas (DEMPSTER; MADAN; CONT, 2016). Os autores também destacam que um contrato a prazo futuro entre um comprador e um vendedor é um acordo no qual o vendedor precisa entregar uma quantidade especificada de mercadoria ao comprador em uma data futura pelo preço combinado anteriormente.

Segundo Kent, Filbeck e Harris (2020), outro modo de investimento em *commodities* é a compra de ações de empresas aos quais seu principal modelo de negócio depende de diferentes *commodities* e de vários setores de mercados físicos. Neste caso, o retorno não depende apenas das *commodities*, mas também da performance do mercado de ações. Kent, Filbeck e Harris (2020) também sugerem o Exchange-Traded Fund (ETF) como forma de investimento. Os ETFs de *commodities* são veículos de investimento para investidores e comerciantes que precisam cobrir riscos ou desejam obter exposição a bens físicos, como produtos agrícolas, recursos energéticos e metais. Tais contratos representam a commodity e acompanham o desempenho de um produto específico preciosos (KENT; FILBECK; HARRIS, 2020).

Carrara e Barros (2018) por sua vez, avaliaram como os choques provenientes dos preços das *commodities*, têm impactado na inflação brasileira e como e com que eficácia a política monetária do país tem reagido. Segundo os autores, a taxa de inflação é influenciada pela expectativa que o mercado forma ao seu respeito e, depois, pelo comportamento dos preços do lado da oferta. Para Carrara e Barros (2018), os desvios da inflação em relação ao esperado se atribuem a variações não previstas de choques de oferta (câmbio e commodities) e do hiato do produto. Já Chisari, Mastronardi e Romero (2019) ao estudarem a vulnerabilidade de três economias latino-americanas (Argentina, Chile e Brasil), relatam que o Brasil seria o menos afetado por quedas nos preços das commodities, porque sua economia é mais diversificada e o setor manufatureiro poderia ser beneficiado pela redução dos preços dos insumos importados. Entretanto, ressaltam que o aumento nos preços de commodities importadas pode ter um impacto negativo na economia do país.

Charles Dow, editor do The Wall Street Journal no final do século XIX, publicou diversos artigos sobre o mercado de ações. Suas teorias se resumem à seis princípios e são conhecidos até hoje como a base da análise técnica, sendo eles (i) os preços descontam tudo, (ii) o mercado tem três tendências, (iii) a tendência primária tem três fases, (iv) o volume deve confirmar a tendência, (v) a tendência precisa ser confirmada por dois índices e (vi) uma tendência é válida até que o mercado indique um sinal definitivo de reversão (ABE, 2009).

Diante desses princípios, Martins (2010) define a análise técnica como uma forma de estudar as cotações passadas de um ativo. Com ela, deve-se encontrar padrões e aplicá-los ao futuro, criando previsões de desempenho como o preço e a duração das operações. Ainda, segundo Martins (2010), para que a análise seja feita, é necessário dividir as cotações em períodos predeterminados, como dias, semanas ou meses. É também muito comum utilizar períodos bem menores, de minutos, como um minuto, dois minutos e assim por diante, normalmente até o máximo de cento e vinte minutos.

Abe (2009) também descreve alguns princípios básicos da análise técnica, seus conceitos, características, padrões de análise, ferramentas e estratégias, uma explanação de gestão de capital e organização de operações. Como indicadores, o autor cita as médias móveis, que são linhas traçadas dentro de um gráfico de preço que se movem sempre que um novo preço é inserido no gráfico, a Média Móvel Convergente e Divergente (MMCD) que mostra a diferença entre duas Médias Móveis Exponenciais (MME), uma rápida (normalmente, de doze períodos) e outra lenta (normalmente, de vinte e seis períodos) e o Índice de Força Relativa (IFR), que mede quantos dias dentro de um período predeterminado foram de alta e quantos foram de baixa, mostrando o resultado em uma escala de 0 a 100%.

Por outro lado, Martins (2010) categorizou como indicadores básicos as médias móveis, o índice de força relativa e a sequência Fibonacci para a criação de retas de suporte e resistência. Conforme o autor, eles são utilizados para a tomada de decisão de compra e venda, onde o preço deve cair para encontrar o suporte ou subir até encontrar a resistência.

Além disso, Martins (2010) também aponta que figuras geométricas podem ser desenhadas em um gráfico a partir das linhas de tendência de alta e de baixa, junto com retas horizontais de suporte e resistência. Tais figuras identificam pontos de compra e venda com projeções de preço-alvo e seu tempo de duração mais provável, dentre elas, foram citadas o triângulo, a Bandeira ou Flâmula, o Retângulo, o Ombro-Cabeça-Ombro (OCO) e o Ombro-Cabeça-Ombro Invertido (OCOI).

Como indicadores avançados, Martins (2010) categorizou o indicador (i) Agulhadas - Didi Index, que indicam quando os preços devem subir ou cair rapidamente; (ii) Bandas de Bollinger, que são a demonstração visual da média e do desvio padrão dos preços de um ativo dentro de seu próprio gráfico, os *gaps* que são intervalos vazios entre os preços dos *candles*, ou seja, patamares de preço em que não houve negócio algum; (iii) Ilhas de Reversão, que ocorrem quando dois *gaps*, um para cima e outro para baixo, ocorrem na mesma faixa de preço, separando o gráfico em duas áreas; (iv) Movimento Direcional que define a tendência atual de um ativo, além de emitir sinais de compra e venda e prever reversões; (v) Estocástico cujo mostra o valor do último preço de fechamento em relação à faixa máximo-mínimo do período analisado e o (vi) Parabólico Stop and Reverse (SAR) que é um sistema de operação o qual baseia-se na premissa de que o investidor sempre estará posicionado no mercado.

Silva (2019) indica que, nos dias atuais, o sucesso financeiro das atividades rurais depende muito do êxito na comercialização das commodities agrícolas. Segundo a autora, o aumento dos custos de produção agrícola observados nos últimos anos, associado à estagnação de produtividade das principais commodities agrícolas têm obrigado os agricultores a realizar um planejamento cada vez mais eficiente da comercialização da produção, fazendo-se necessário compreender os diferentes mecanismos de comercialização das commodities agrícolas disponíveis e traçar estratégias para se proteger das oscilações do mercado e para garantir maior lucratividade da atividade agrícola.

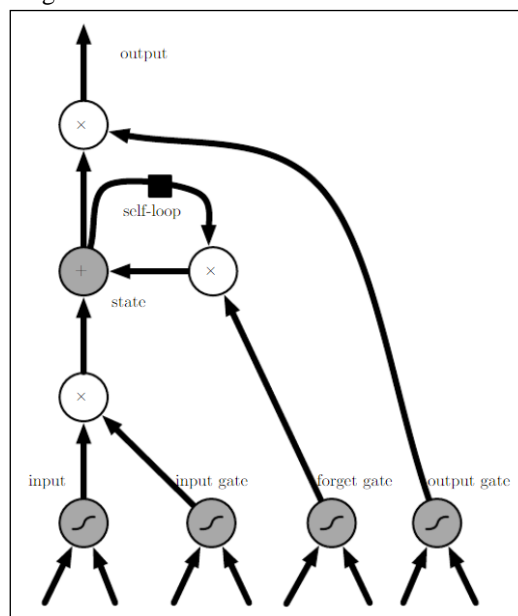
2.2 APRENDIZADO DE MÁQUINA E AS REDES NEURAIS LSTM

Segundo Mitchell (1997), para a construção de um sistema de aprendizado, é necessário definir a tarefa a ser realizada, a medida de desempenho e a experiência de treino. Ainda segundo o autor, deve ser extraído o tipo de conhecimento a ser aprendido, ter uma representação desse conhecimento e um mecanismo de aprendizado. Mitchell (1997) destaca que uma medida de desempenho frequentemente utilizada em modelos de regressão é o *Mean Squared Error* (MSE), que consiste na média dos erros de predição quadráticos em todas as instâncias do conjunto de dados, no qual o erro de previsão é a diferença entre o valor verdadeiro e o valor previsto para uma instância.

Goodfellow, Bengio e Courville (2016) apontam três categorias de algoritmos inteligentes para a construção de sistemas de aprendizado. Na primeira, encontram-se os sistemas baseados em regras, no qual a partir de uma entrada, é realizado um processamento procedural, gerando uma saída. Na segunda, encontram-se os algoritmos clássicos de aprendizado de máquina, nesta, a partir de uma entrada, é realizado o processamento manual das características e produzido um mapeamento a partir das mesmas, e ao completar o mapeamento, gera-se a saída. Por fim, encontram-se os modelos de representação, estes processam as características a partir do aprendizado dos dados, mapeiam as características e produzem o resultado. Os autores também adicionam uma subcategoria aos modelos de representação, chamada de Aprendizado Profundo, que além de conter todas as características de um modelo de representação, possuem um mecanismo onde características mais abstratas são extraídas a partir de camadas adicionais inseridas no algoritmo. Tal subdivisão incluem as redes neurais artificiais.

Redes neurais que utilizam a arquitetura de células *Long Short Term Memory* (LSTM) possuem mecanismos que permitem que a rede neural acumule informações sobre uma longa duração de tempo (KOSTADINOV, 2017). De acordo com Goodfellow, Bengio e Courville (2016), estas redes neurais possuem células conectadas recorrentemente umas às outras e seus valores de entrada podem ser acumulados em uma unidade de estado (*state unit*), a qual possui um *loop* linear próprio, com peso controlado pelo portão de esquecimento (*forget gate*). Goodfellow, Bengio e Courville (2016) também ressaltam que a saída da célula pode ser desligada pela porta de saída (*output gate*) e que todas as unidades de passagem têm uma não-linearidade sigmoide, enquanto a unidade de entrada pode ter qualquer função não linear de compressão. Ainda segundo os autores, a unidade de estado também pode ser usada como uma entrada extra para as outras portas da célula (*gating units*). Na Figura 1, é ilustrado o diagrama de blocos de uma célula LSTM, onde os nós escuros indicam o um único intervalo de tempo (*time step*).

Figura 1– Diagrama de blocos de uma célula da rede recorrente LSTM



Fonte: Goodfellow, Bengio e Courville (2016, p. 398).

Srivastava *et al.* (2014) indicam que com dados de treinamento limitados, muitas relações complexas entre as entradas e as saídas podem ser resultado de ruído de amostragem, existindo no conjunto de treinamento mas não no conjunto de teste, que por consequência gera o problema do *overfitting*, no qual a rede neural generaliza o conjunto de treino, porém falha ao realizar previsões para os dados de teste. Para resolver tal problema, Srivastava *et al.* (2014) propôs uma técnica nomeada de *dropout*, que além reduzir significativamente o *overfitting*, fornece uma maneira de combinar diferentes arquiteturas de redes neurais. Ele consiste em desligar temporariamente unidades da rede neural na fase de treinamento, junto com todas suas conexões de entrada e saída, de maneira aleatória, com uma probabilidade pré definida de desligamento. Já na fase de teste as unidades estão sempre presentes, entretanto os pesos são multiplicados pela probabilidade definida.

Por fim, redes neurais podem ter sua etapa de treinamento otimizada. O algoritmo Adam de Kingma e Ba (2014), que é derivativo de Estimativa de Momento Adaptativo (*Deadaptive Moment Estimation - DME*), pode ser utilizado na substituição do clássico *Stochastic Gradient Descent* (SGD) para a atualização iterativa dos pesos da rede. Este algoritmo usa uma taxa de aprendizagem para cada peso da rede e adapta-se separadamente enquanto o treinamento ocorre. Suas principais vantagens, segundo os autores, se dão por ser computacionalmente eficiente, apropriado para problemas com gradientes com ruído ou esparsos, bem adequado para problemas com grande número de dados ou parâmetros, e por fim, pelo baixo custo de memória.

3 TRABALHOS CORRELATOS

Esta seção apresenta na forma de quadros, três trabalhos correlatos cujo objetivos são semelhantes ao do tema proposto. O trabalho de Wang, Liu e Wu (2019) investigou e demonstrou a previsibilidade das mudanças nos preços das commodities utilizando indicadores técnicos em oito classes de *commodities*. Zhang *et al.* (2020) propuseram um *framework* de seleção de modelos que inclui recursos de séries temporais e horizontes de previsão para a predição de preços de commodities agrícolas. Por fim, o trabalho de Fang *et al.* (2020) utilizou a decomposição do modo empírico e um conjunto algoritmos de aprendizado de máquina para realizar a estimativa de preços de commodities agrícolas.

Quadro 1 – Forecasting commodity prices out-of-sample: Can technical indicators help?

Referência	Wang, Liu e Wu (2019)
Objetivos	Investigar a habilidade preditiva da análise técnica sobre os preços de commodities.
Principais funcionalidades	Foi desenvolvido um modelo de predição utilizando diversos indicadores de análise técnica, no qual, para fins de comparação, o modelo foi utilizado com variáveis macroeconômicas, evidenciando que, por mais que indicadores técnicos tenham recebido menos atenção da literatura, eles possuem maior habilidade preditiva.

Ferramentas de desenvolvimento	Foram utilizados dados mensais dos preços de oito índices de commodities (Energia, Não-energia, Agricultura, Bebidas, Alimentos, Matéria-prima, Metais & Minerais e Metais preciosos) obtidos do site do Banco Mundial e, que continham amostras de janeiro de 1982 à dezembro de 2017. Wang, Liu e Wu (2019) também utilizaram oito variáveis macroeconômicas (relação de preço de dividendos, taxa de letras do Tesouro, rendimento a longo prazo, taxas de câmbio, excesso de retorno de estoque, inflação, produção industrial nos EUA e atividade econômica real global conforme procurada pelo índice de Kilian), sendo obtidas a partir da homepage de Amit Goyal, do site do Banco da Reserva Federal de Saint Louis e da homepage de Lutz Kilian.
Resultados e conclusões	Comparando os valores preditos com o valor real das commodities, Wang, Liu e Wu (2019) notaram que ao utilizar indicadores de análise técnica, a porcentagem fora da amostra (R^2) indicou significados positivos para ambas as oito classes de commodities, enquanto ao utilizar variáveis macroeconômicas, o resultado demonstrou uma boa significância para apenas seis categorias, indicando que, indicadores técnicos possuem um maior poder preditivo que indicadores macroeconômicos. Contudo, Wang, Liu e Wu (2019) relataram que suas descobertas confirmam o significado econômico da previsibilidade do preço das commodities e que, os indicadores técnicos têm desempenho melhor que os indicadores econômicos para alocação de portfólio.

Fonte: elaborado pelo autor.

Quadro 2 – Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons

Referência	Zhang <i>et al.</i> (2020)
Objetivos	Criar um <i>framework</i> de seleção de modelos que inclui recursos de séries temporais e horizontes de previsão visando prever preços de <i>commodities</i> agrícolas.
Principais funcionalidades	Foi selecionado um modelo preditor a partir de um determinado conjunto de entradas. Para isto, definiu-se três módulos principais, nomeados, extração de características, seleção de características e classificação.
Ferramentas de desenvolvimento	Zhang <i>et al.</i> (2020) utilizaram uma base composta por 522 séries de preços de <i>commodities</i> agrícolas mensais, cobrindo as <i>commodities</i> leiteiro, carne de porco, farelo de trigo, pintinho, milho, grãos para porcos, carne de carneiro, carne de frango, frango novo, lã, porco, carne de bovino, comida suína, ovo, feijão, farinha de peixe, frango vivo, frango simples e camada de alimentação.
Resultados e conclusões	Comparando os resultados dos três modelos de previsão (RNA, SVR e ELM), Zhang <i>et al.</i> (2020) observaram que a RNA é o modelo mais eficaz, pois possui o menor erro médio de previsão. Já em relação aos dois classificadores utilizados no estudo (RF e SVM), verificaram que o erro de previsão do SVM é, em média, maior que o do RF em todos os casos, o que torna o RF, neste caso, um algoritmo mais adequado para a realização da tarefa de classificação. No que diz respeito a estrutura de seleção de modelos, Zhang <i>et al.</i> (2020) observaram que o erro médio de previsão do MSN-RF é de 8,6673 em comparação com 8,6744 da RNA. Segundo os autores, esse resultado demonstra a superioridade da estrutura de seleção de modelos, que podendo efetivamente resultar em um erro de predição menor.

Fonte: elaborado pelo autor.

Quadro 3 – Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices

Referência	Fang <i>et al.</i> (2020)
Objetivos	Realizar a regressão de preços de <i>commodities</i> a partir de técnicas de Decomposição do Modo Empírico.
Principais funcionalidades	Predizer valores para os preços futuros das <i>commodities</i> agrícolas em um intervalo de três dias à frente.
Ferramentas de desenvolvimento	Foram utilizados dados coletados de 27 de novembro de 2014 à 11 de outubro de 2017 do Wind database, sendo utilizadas as categorias: vegetable meal, soybean meal, stalked rice, Strong wheat, Zheng Cotton e early indica rice.
Resultados e conclusões	Fang <i>et al.</i> (2020) demonstram que em média o DME/RNA tem uma melhor performance para prever valores de um passo à frente, visto que obteve-se um RRMSE médio de 0,7067, valor superior aos outros algoritmos. Por outro lado, para a predição de três passos à frente, destaca-se o algoritmo DME/ARIMA, que obteve uma performance de 0,6657.

	Fang <i>et al.</i> (2020) indicam que o desempenho da previsão dos modelos combinados foi superior ao dos modelos individuais e que a partir de seus resultados é possível concluir que com o aumento horizontal nos dias de predição, a superioridade dos modelos combinados usando a decomposição do modo empírico (DME) torna-se mais evidente. Além disso, os autores também ressaltam que o SVM e a RNA são mais adequadas para a previsão de componentes de alta frequência em comparação aos modelos ARIMA.
--	--

Fonte: elaborado pelo autor.

4 DESCRIÇÃO E RESULTADOS DA APLICAÇÃO

Este capítulo descreve os aspectos mais relevantes relacionados ao desenvolvimento da aplicação, apresentando os Requisitos Funcionais (RF) no Quadro 4 e os Requisitos Não Funcionais (RNF) no Quadro 5. Também apresenta detalhes sobre a implementação, contemplando a descrição dos dados, o pré-processamento realizado, as arquiteturas de redes neurais utilizadas e as etapas seguidas no desenvolvimento da aplicação.

Quadro 1 – Requisitos Funcionais

RF01 - permitir o carregamento de dados de commodities de um mercado financeiro internacional e de um mercado nacional
RF02 - processar indicadores técnicos referentes aos dados financeiros das commodities
RF03 - prever o preço das commodities selecionadas em uma perspectiva de 1 e 3 dias a frente
RF04 - exibir o desempenho dos modelos através do <i>mean squared error</i> (MSE)
RF05 - exibir as performances dos modelos, preços reais e preditos em um <i>dashboard</i>

Fonte: elaborado pelo autor.

Quadro 5 – Requisitos Não Funcionais

RNF01 - ser desenvolvido utilizando a linguagem de programação Python
RNF02 - utilizar as bibliotecas Pandas e Scikit-learn para o pré-processamento de dados
RNF03 - utilizar a biblioteca Keras para desenvolvimento e treinamento dos modelos
RNF04 - utilizar a plataforma Jupyter Notebook para exibição dos resultados
RNF05 - utilizar a biblioteca TA-Lib para o cálculo dos indicadores técnicos
RNF06 - utilizar a arquitetura de rede neural <i>long short term memory</i> no desenvolvimento dos modelos preditores

Fonte: elaborado pelo autor.

Na aplicação, utilizou-se as bases de dados disponibilizadas por (i) Yahoo! – Verizon Media (2020) – indicadores internacionais e (ii) Universidade De São Paulo (2020) – indicadores nacionais, tendo informações financeiras de cinco *commodities* agrícolas, sendo elas o açúcar, algodão, milho, soja e o trigo, com frequência diária no intervalo de 03 de janeiro de 2005 à 31 de dezembro de 2019. As variáveis contidas nas bases de dados disponibilizadas por (i) são os valores de abertura, máximo, mínimo, fechamento, ajuste de fechamento e volume, enquanto na base de dados disponibilizada por (ii) é contido o preço do dia em reais e em dólares, no qual foi selecionado para este projeto os valores em reais. O Quadro 6 apresenta a relação entre as bases de dados, indicando seus respectivos códigos para obtenção.

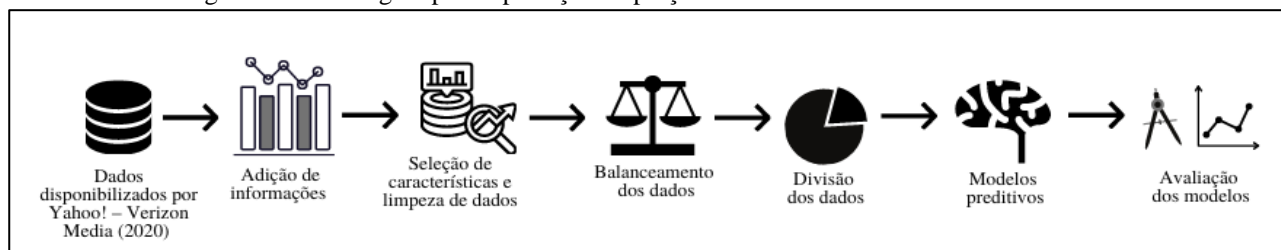
Quadro 2 – Relacionamento entre as bases de dados

Commodity	Código – Yahoo! Verizon Media (2020)	Código - Universidade De São Paulo (2020)
Açúcar	SB=F	Indicador do açúcar cristal CEPEAESALQ - São Paulo
Algodão	CT=F	Indicador do algodão em pluma CEPEAESALQ - prazo de 8 dias
Milho	ZC=F	Indicador do milho ESALQBM&FBOVESPA
Soja	ZS=F	Indicador da soja CEPEAESALQ - Paraná
Trigo	KW=F	Preço médio do trigo CEPEAESALQ - Paraná

Fonte: elaborado pelo autor.

Visando adicionar informações à base de dados nacionais, foram criados modelos preditores para o mercado internacional de commodities que utilizaram como variáveis de entrada os valores de máximo, mínimo e volume, além de indicadores de análise técnica. Suas predições foram incorporadas à base de dados nacionais. A Figura 2 demonstra o fluxo seguido para realizar a predição de preços futuros de commodities internacionais.

Figura 2 – Abordagem para a predição de preços futuros de *commodities* internacionais



Fonte: elaborado pelo autor.

Pode-se notar através da Figura 2 que após obter a base de dados, iniciou-se o processo de adição de informações. Essa etapa consiste no cálculo dos indicadores técnicos a partir da biblioteca de análise técnica TA-Lib, uma biblioteca de código aberto disponível para as linguagens de programação C/C++, Java, Perl, Python e .NET a qual é amplamente utilizada em software que exigem análise técnica de dados de mercados financeiros, pois possibilita o cálculo de indicadores de análise técnica de forma automatizada (TICTACTEC LLC. *et al.*, 2007). Foram adicionados à base de dados disponibilizada por Yahoo! Verizon Media (2020), 3 indicadores de volume, 30 indicadores de momento, 3 indicadores de volatilidade e 7 detectores de ciclo, totalizando 43 indicadores técnicos. A Tabela 1 apresenta os indicadores técnicos calculados, classificando-os pelo seu tipo.

Tabela 1 – Indicadores técnicos calculados sobre a base de dados internacionais, classificados pelo seu tipo.

Tipo	Indicadores calculados (Yahoo! Verizon Media - 2020)
Detectores de ciclo	HT DCPERIOD, HT DCPHASE, INPHASE, QUADRATURE, SINE, LEADSINE, HT TRENDMODE
Indicadores de momento	ADX, ADXR, APO, AROON, AROONOSC, BOP, CCI, CMO, DX, MACD, MACDEXT, MACDFIX, MIF, MINUS DI, MINUS DM, MOM, PLUS DI, PLUS DM, PPO, ROC, ROCP, ROCR, ROCR100, RSI, STOCH, STOCHF, STOCHRSI, TRIX, ULTOSC, WILLR
Indicadores de volatilidade	ATR, NATR, TRANGE
Indicadores de volume	AD, ADOSC, OBV

Fonte: elaborado pelo autor.

Realizado os cálculos dos indicadores técnicos, foram selecionadas as características mais relevantes para a predição de dados de cada *commodity*. Para isso, foi calculado a correlação de Pearson entre cada indicador técnico com os dados de fechamento do dia. Esta correlação indica a força da associação linear entre duas variáveis e pode ser calculada através da Equação 1.

Equação 1 – Cálculo da correlação de Person

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Fonte: elaborado pelo autor.

Onde r é o coeficiente de correlação de Pearson, x_i os valores da variável x em uma amostra, \bar{x} a média dos valores da variável x , y_i os valores da variável y e \bar{y} a média dos valores da variável y .

Nesta etapa, foram mantidos os indicadores cujo módulo de correlação fosse superior a 0,15, visto que ao utilizar tal valor, manteve-se uma quantidade considerável de indicadores técnicos, selecionando aqueles que mais se destacaram. A Tabela 2 descreve os coeficientes encontrados para cinco indicadores técnicos referente aos valores de fechamento de mercado das *commodities* para a base de dados disponibilizada por Yahoo! Verizon Media (2020). Com isso, foi realizado a limpeza dos dados, a qual excluiu-se as primeiras linhas do conjunto de dados que continham o valor nulo devido ao pré-processamento dos indicadores técnicos.

Tabela 2 – Coeficiente de correlação de Pearson sobre os dados internacionais de mercado.

Indicador / Commodity	Açúcar	Algodão	Milho	Soja	Trigo
NATR	0,31008719	0,35868731	0,31442310	0,28694566	0,11916946
OBV	0,45523242	0,16162160	0,23452564	0,30119409	0,59871290
MINUS DM	0,58859266	0,63013850	0,50869227	0,38265851	0,59956594
TRANGE	0,62805229	0,69649601	0,56482108	0,47438842	0,47674641
PLUS DM	0,73745874	0,83559670	0,67304095	0,61342416	0,75163947

Fonte: elaborado pelo autor.

Observando a Tabela 2, nota-se que o indicador NATR obteve um índice de correlação menor que 0,15 para a *commodity* trigo, sendo excluído da base de dados. A tabela completa contendo todos os indicadores técnicos e suas respectivas correlações pode ser encontrada na Tabela 7 do Apêndice A.

Após selecionar os indicadores técnicos com correlação superior a 0,15 e realizado a limpeza dos dados, empregou-se a técnica de normalização de dados. Esta etapa definiu um intervalo de valores para os dados visando reduzir a diferença entre o menor e o maior valor do conjunto. A fórmula padrão para a normalização é apresentada na Equação 2.

Equação 2 – Fórmula padrão da normalização

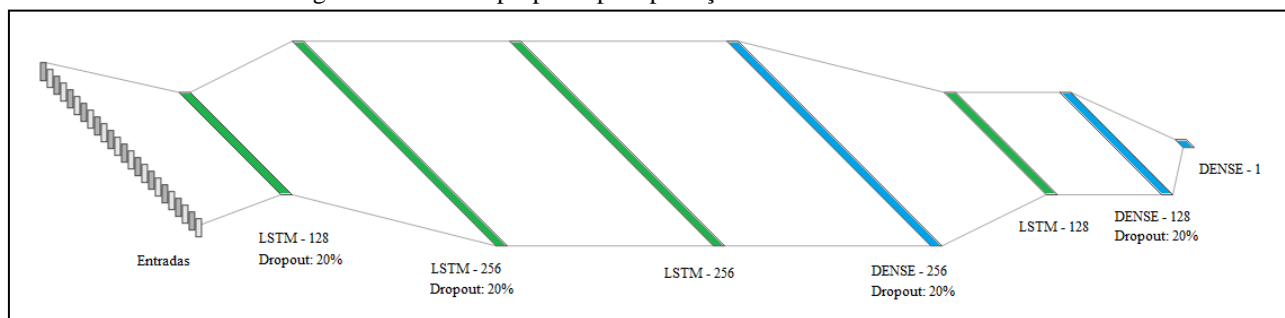
$$y = \frac{(x - x_{min}) * (d_{max} - d_{min})}{x_{max} - x_{min}} + d_{min}$$

Fonte: elaborado pelo autor.

Onde, y representa o valor normalizado, x o valor a ser normalizado, x_{max} se refere ao valor máximo a ser normalizado, x_{min} ao valor mínimo a ser normalizado, d_{max} ao valor máximo desejado e d_{min} ao valor mínimo desejado. Neste projeto, foi utilizado 1 (um) como valor máximo desejado e 0 (zero) para o valor mínimo desejado.

Após a etapa de normalização de dados, o conjunto disponibilizado por Yahoo! – Verizon Media (2020) foi dividido na proporção de 20% para treinamento e 80% para teste. Com isso, tornou-se possível a construção e treinamento dos modelos preditivos para as *commodities* no âmbito do mercado internacional. A arquitetura destes modelos consistiu de uma camada LSTM com 128 unidades e *dropout* de 20%, seguido de uma camada LSTM com 256 unidades e *dropout* de 20%. Outra camada LSTM com 256 unidades, uma camada totalmente conectada (*Dense*) com 256 neurônios e *dropout* de 20%, uma camada LSTM de 128 unidades, uma camada totalmente conectada de 128 unidades com *dropout* de 20%. E, por fim, uma camada totalmente conectada com um neurônio e a função de ativação *Rectified Linear Unit* (ReLU). Esta arquitetura pode ser visualizada graficamente através da Figura 3.

Figura 3 – Modelo proposto para predições no âmbito internacional



Fonte: elaborado pelo autor.

Concluído a etapa de treinamento, os modelos preditores para o mercado internacional foram avaliados utilizando como métrica de desempenho o *Mean Squared Error* (MSE), cujo resultados para todos os *commodities* utilizados, com predições para 1 e 3 dias à frente são descritos na Tabela 3. Por fim, tais modelos foram utilizados para gerar uma base de dados de predições do mercado internacional de *commodities* sobre o conjunto de teste.

Tabela 3 – MSE dos modelos de redes neurais referente ao mercado internacional.

Commodity	MSE – Fase de treino: 1 dia à frente	MSE – Fase de teste: 1 dia à frente	MSE – Fase de treino: 3 dias à frente	MSE – Fase de teste: 3 dias à frente
Açúcar	0,00040	0,00228	0,00073	0,00172
Algodão	0,00062	0,00011	0,00139	0,00022
Milho	0,00039	0,00062	0,00114	0,00118
Soja	0,00062	0,00113	0,00165	0,00308
Trigo	0,00048	0,00056	0,00108	0,00153

Fonte: elaborado pelo autor.

O MSE indica o quão próximo os valores de duas curvas se encontram, sendo zero o menor valor possível, o qual indica total sobreposição, e um o máximo valor possível, o qual indica a pior sobreposição. Pode-se verificar através da Tabela 3, que para todas as *commodities* o valor do MSE ficou abaixo de 0,005, indicando que as curvas de predição foram próximas as curvas de preços reais, tendo destaque o algodão, que obteve os menores valores do MSE na fase de teste, para ambas as predições de 1 e 3 dias à frente.

Para estabelecer os indicadores do mercado nacional, foram realizados processos semelhantes sobre a base de dados disponibilizada por Universidade De São Paulo (2020). A Figura 4 demonstra o fluxo seguido para realizar as predições sobre o mercado nacional de *commodities*.

Figura 4 – Abordagem para a predição de preços futuros de *commodities* nacionais



Fonte: elaborado pelo autor

Para a base de dados do mercado nacional, foram adicionados 13 indicadores de momento e 7 detectores de ciclo, totalizando 20 indicadores técnicos. A diferença na quantidade de indicadores técnicos se deu devido a quantidade de características disponibilizadas por cada base de dados. A Tabela 4 apresenta os indicadores técnicos utilizados, classificando-os pelo seu tipo.

Tabela 4 – Indicadores técnicos calculados sobre a base de dados nacionais, classificados pelo seu tipo.

Indicadores	Nacionais (Universidade De São Paulo)
Detectores de ciclo	HT DCPERIOD, HT DCPHASE, INPHASE, QUADRATURE, SINE, LEADSINE, HT TRENDMODE
Indicadores de momento	APO, CMO, MACD, MACDEXT, MACDFIX, MOM, PPO, ROC, ROCP, ROCR100, RSI, STOCHRSI, TRIX
Indicadores de volatilidade	Não se aplica
Indicadores de volume	Não se aplica

Fonte: elaborado pelo autor.

Tais indicadores também passaram pelo processo de seleção, tendo como critério o coeficiente de Pearson. A Tabela 5 demonstra os coeficientes de correlação para 5 indicadores técnicos referentes a base de dados disponibilizada por Universidade De São Paulo (2020).

Tabela 5 – Coeficiente de correlação de Pearson sobre os dados nacionais de mercado.

Indicador / Commodity	Açúcar	Algodão	Milho	Soja	Trigo
RSI	0,18509653	0,08100763	0,30984300	0,30517003	0,16748187
APO	0,19549568	0,16702112	0,26767576	0,28277952	0,23069744
MACDFIX	0,26530357	0,21047608	0,34258186	0,38902278	0,31027552
HT DCPERIOD	0,27262300	0,10624508	0,27997955	0,25604055	- 0,02854838
TRIX	0,43366235	0,21416214	0,46843084	0,54283840	0,47789509

Fonte: elaborado pelo autor.

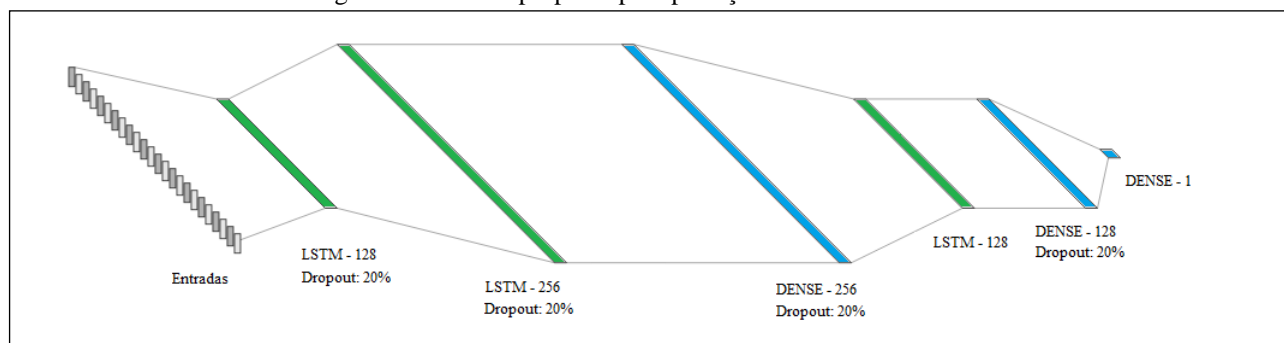
A partir da Tabela 5, pode-se verificar que os valores de correlação para os indicadores técnicos RSI, para a *commodity* algodão e HT DCPERIOD, para as *commodities* algodão e trigo não atingiram um índice superior a 0,15, sendo excluídos da base de dados nacionais para as respectivas *commodities*. A tabela completa contendo todos os indicadores técnicos e suas respectivas correlações encontram-se na Tabela 8 do Apêndice A. Por fim, tais dados também passaram pelo processo de normalização, sendo escalados para o intervalo fechado de 0 à 1.

Com os dados da base da Universidade De São Paulo (2020) normalizados, foram emparelhados com os dados preditos para o mercado internacional de *commodities*. Com isso, formou-se o conjunto de entrada para os modelos preditores do mercado nacional. Estes dados por sua vez foram novamente normalizados entre 0 e 1, para o ajuste das predições e subdivididos em dois conjuntos, sendo 80% dos dados para o treinamento e 20% para teste.

A partir disso, foram criados os modelos preditores para o mercado nacional de *commodities*, estes, compostos por uma camada LSTM com 128 unidades e *dropout* de 20%, seguido de uma camada LSTM com 256 unidades, uma camada totalmente conectada com 256 neurônios e *dropout* de 20%, uma camada LSTM de 128 unidades, uma camada

totalmente conectada de 128 unidades com *dropout* de 20% e por fim, uma camada totalmente conectada com um neurônio e função de ativação ReLU. O modelo descrito pode ser visualizado graficamente na 5.

Figura 5 – Modelo proposto para predições no âmbito nacional



Fonte: elaborado pelo autor.

Este modelo por sua vez, também passou pelo processo de avaliação, utilizando também como métrica de desempenho o MSE, cujo resultados para todas as *commodities* utilizadas, com predições para 1 e 3 dias à frente podem ser encontrados na Tabela 6. Com isso, foi possível realizar as predições para o mercado nacional de commodities, cumprindo com o objetivo proposto.

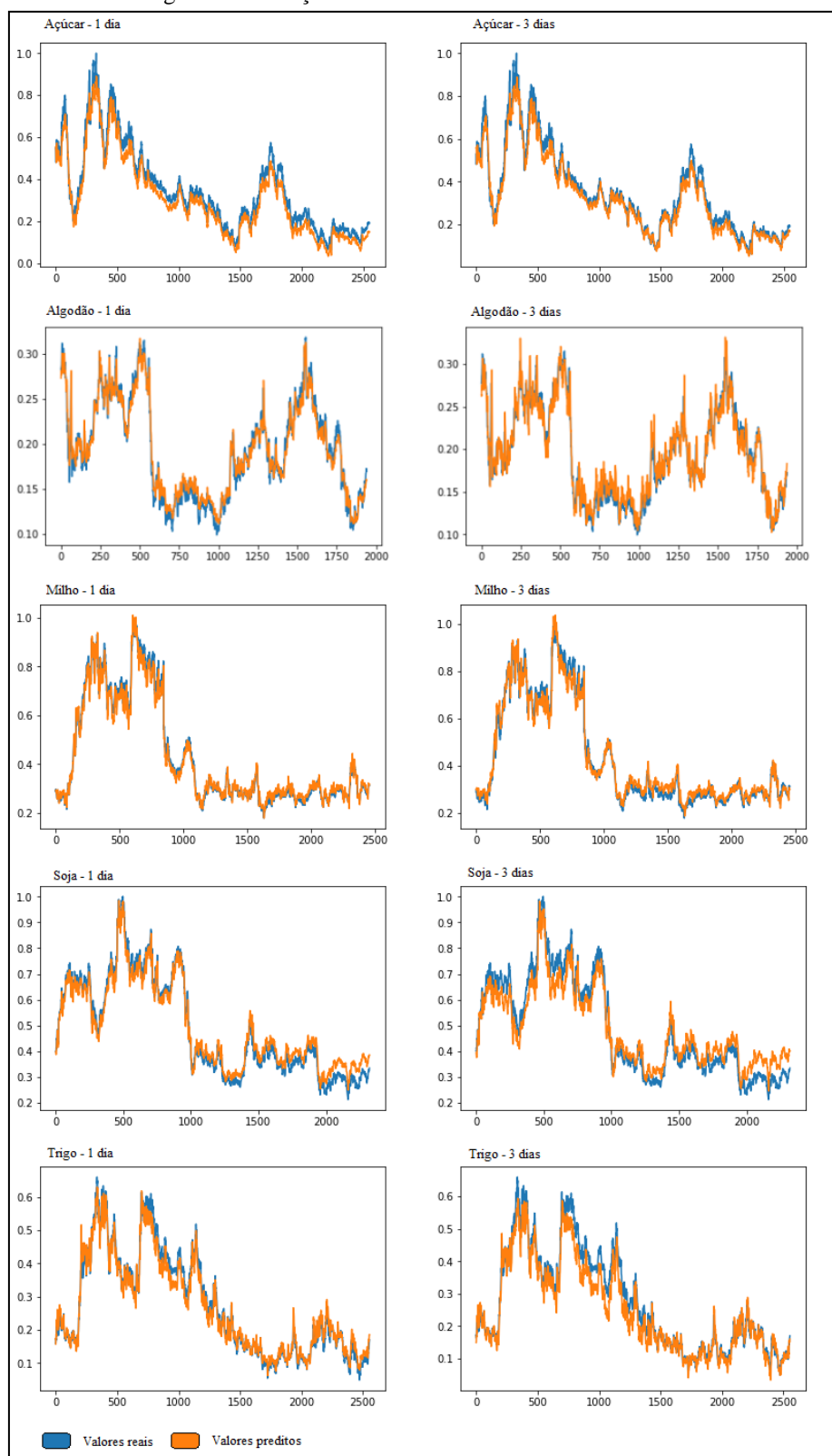
Tabela 6 – MSE dos modelos de redes neurais referente ao mercado nacional.

Commodity	MSE – Fase de treino: 1 dia à frente	MSE – Fase de teste: 1 dia à frente	MSE – Fase de treino: 3 dias à frente	MSE – Fase de teste: 3 dias à frente
Açúcar	0,00038	0, 00019	0,00027	0,00017
Algodão	0,00023	0, 00010	0,00025	0,00015
Milho	0,00082	0, 00027	0,00030	0,00030
Soja	0,00047	0, 00037	0,00045	0,00042
Trigo	0,00028	0, 00016	0,00021	0,00017

Fonte: elaborado pelo autor.

Contudo, a partir dos resultados descritos nas Tabelas 5 e 6, é notável o alto poder preditivo dos modelos desenvolvidos, tendo destaque para a *commodity* algodão, que obteve os menores valores de MSE para ambas as predições referentes aos dados de teste. Na Figura 5 são exibidos os gráficos das predições encontradas pelos modelos de redes neurais no âmbito internacional em comparação com os dados reais no conjunto de teste, e na Figura 6, são exibidos os gráficos das predições encontradas pelos modelos no âmbito nacional, em comparação com os valores reais, novamente no conjunto de teste.

Figura 5 – Predições dos modelos no âmbito internacional

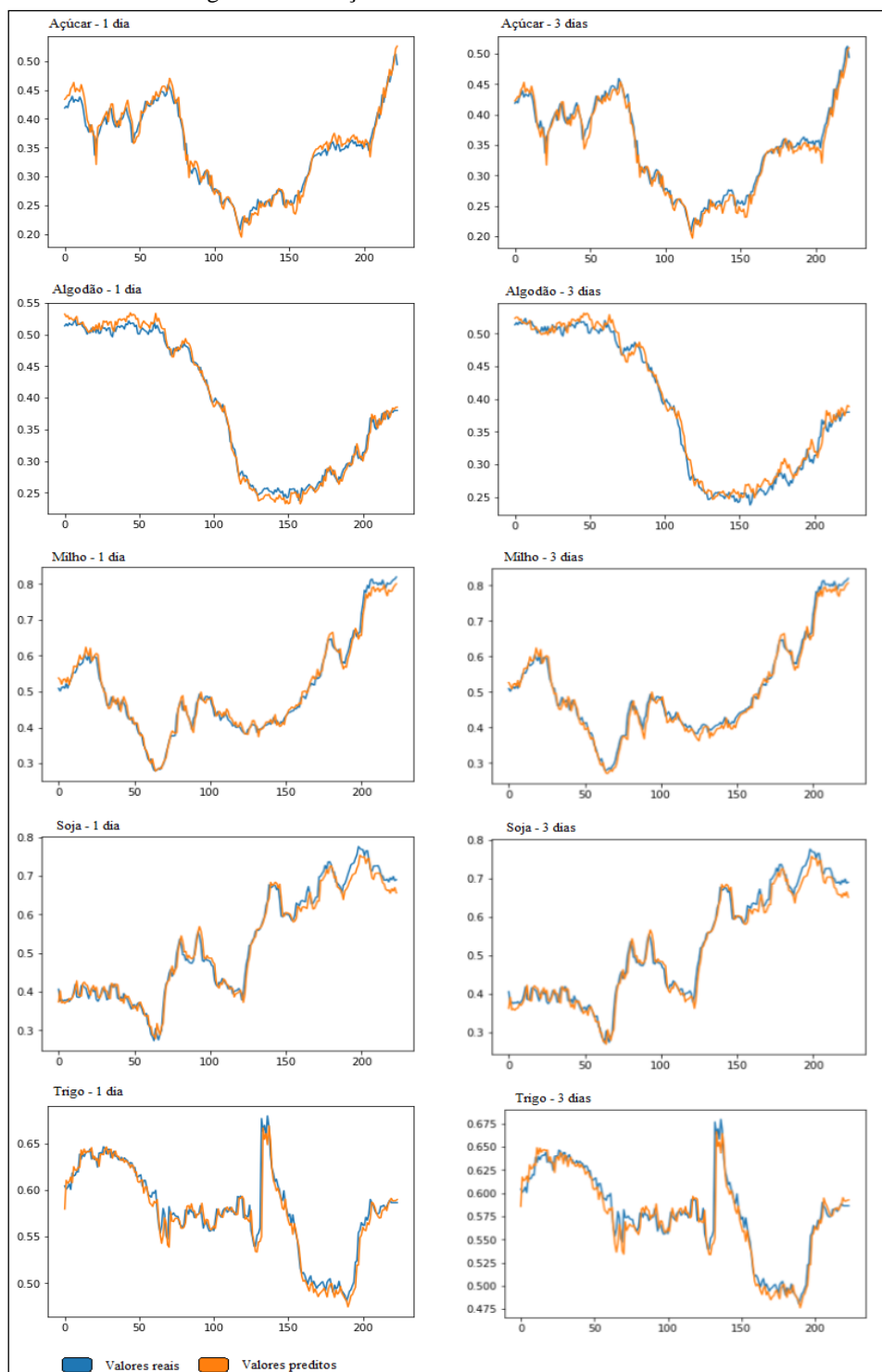


Fonte: elaborado pelo autor.

A partir da Figura 5 pode-se observar que, apesar do modelo ter atingido boas predições, em algumas situações a curva de predições se desloca em relação aos dados reais, em destaque para a *commodity* soja, que apresentou as piores predições para 3 dias a frente. Esta situação confirma a avaliação obtida pela métrica de desempenho *MSE*, no qual a soja apresentou o pior resultado, com um valor de 0,0308 para predições de 3 dias a frente. Destaca-se que, para os dados do mercado internacional, apenas 20% do conjunto de dados foi utilizado para treino, o que pode ter influenciado tais desvios na curva de predição. Por outro lado, na Figura 6 as curvas de predições possuem um deslocamento sucinto em

relação aos dados reais, novamente confirmando os dados apresentados pela métrica de desempenho e indicando que a metodologia apresentada neste trabalho é apropriada para realizar a predição de commodities agrícolas.

Figura 6 – Predições dos modelos no âmbito nacional



Fonte: elaborado pelo autor

5 CONCLUSÕES

A predição de preços de commodities agrícolas é uma importante tarefa econômica e impacta em diversas classes sociais, sendo elas atingidas direta ou indiretamente. Visando obter predições precisas nesta área, este trabalho apresentou uma abordagem para a predição de preços futuros de *commodities* agrícolas utilizando indicadores de análise técnica e aprendizado de máquina. Para isso, utilizou-se duas bases de dados, as quais continham informações financeiras sobre as *commodities* açúcar, algodão, milho, soja e trigo, com dados frequência diária no intervalo de 03 de janeiro de 2005 à 31

de dezembro de 2019. Tais bases de dados passaram pelos processos de limpeza, pré-processamento e subdivisão de dados para então serem utilizadas no treinamento e avaliação dos modelos desenvolvidos.

O projeto foi desenvolvido a partir da linguagem de programação Python, sendo utilizadas as bibliotecas Pandas e Scikit-learn para o pré-processamento de dados, TA-Lib para o cálculo dos indicadores técnicos, Keras para desenvolvimento e treinamento dos modelos além da plataforma Jupyter Notebook para exibição dos resultados. Tais tecnologias se mostraram efetivas, cumprindo com seu propósito.

No decorrer do projeto foram adicionados à base de dados internacionais, 43 indicadores técnicos. Já na base de dados nacionais, foram adicionados 20 indicadores técnicos. Tais indicadores passaram por um processo seletivo, o qual manteve nas bases de dados aqueles cujo módulo de correlação de Pearson fosse superior a 0,15 em relação aos dados de fechamento do mercado. Com isso, os dados foram normalizados, sendo escalados para o intervalo fechado de 0 à 1. Concluído estes processos, dividiu-se a base de dados internacionais em 20% para treinamento e 80% para teste. Utilizou-se então esta base no treinamento dos modelos preditores para o mercado internacional de commodities. Estes modelos por sua vez, geraram previsões referentes aos dados de teste, sendo acopladas a base de dados das commodities nacionais. A partir disso, normalizou-se novamente a base de dados nacionais que em seguida foi dividida em 80% para treinamento e 20% para teste. Por fim, foram criados os modelos preditivos para esta base de dados, os quais obtiveram resultados entre 0,00009 e 0,00037 MSE nos dados de teste para 1 dia à frente e de 0,00015 a 0,00041 MSE nos dados de teste para 3 dias à frente, indicando uma boa performance de previsão para todas as commodities avaliadas, concluindo o objetivo do projeto.

Em relação aos trabalhos correlatos, este projeto apresentou uma abordagem a qual unificou a análise técnica, objeto de estudo de Wang, Liu e Wu (2019), com algoritmos de aprendizado de máquina, objeto de estudo de Zhang *et al.* (2020) e Fang *et al.* (2020). Além disso, este projeto destacou-se pelo uso de redes neurais com arquitetura *LSTM* e pelo uso de previsões sobre dados de mercados internacionais para realizar previsões no âmbito do mercado nacional de commodities. Ainda em relação aos trabalhos correlatos, destaca-se que apenas o trabalho de Wang, Liu e Wu (2019) fez o uso de indicadores técnicos, o qual não realizou previsões a curto prazo. Com isso, pode-se inferir que este trabalho contribuiu com a literatura ao apresentar um estudo que utilizou algoritmos de aprendizado de máquina em combinação com indicadores de análise técnica para a previsão de séries temporais à curto prazo, fazendo o uso de dados de mercados internacionais e nacionais para a previsão do preço das commodities no âmbito do mercado brasileiro.

Por fim, destaca-se que este trabalho se limitou a apenas cinco commodities, contudo, sugere-se como extensões: (i) a utilização desta abordagem para prever commodities não agrícolas, como o ouro e o petróleo, (ii) a utilização e aprimoramento da metodologia desenvolvida para previsões em diferentes intervalos de tempo, com frequência superior e inferior ao apresentado neste trabalho, (iii) o aprimoramento dos modelos desenvolvidos através da inserção de outras técnicas de aprendizado de máquina e (iv) o desenvolvimento de um software que utilize a abordagem proposta para realizar previsões sobre o mercado futuro de commodities.

REFERÊNCIAS

- ABE, Marcos. **Manual de análise técnica: Essência e estratégias avançadas**. 1. ed. rev. São Paulo: Novatec Editora Ltda, 2009. 256 p. v. 6. ISBN 978-85-7522-702-2.
- BOROVKOVA, Svetlana; TSIMAS, Ioannis. **An ensemble of LSTM neural networks for high-frequency stock market classification**. Journal of Forecasting, Wiley online library, ed. 38, p. 600– 619, 21 mar. 2019. Disponível em: <https://onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Ffor.2585>. Acesso em: 19 abr. 2020.
- CARRARA, Aniela Fagundes; BARROS, Geraldo S.A de Camargo. **Choques de oferta e política monetária na economia brasileira: Uma análise do impacto dos preços das commodities na inflação entre 2002 e 2014**. Nova Economia, Piracicaba - SP, Brasil., v. 29, n. 3, 24 ago. 2018. DOI <https://doi.org/10.1590/0103-6351/4070>. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-3512019000300757&tlng=pt. Acesso em: 17 jun. 2020.
- CHISARI, Omar O.; MASTRONARDI, Leonardo J.; ROMERO, Carlos A. **Commodities prices and critical parameters for macroeconomic performance: a CGE analysis for Argentina, Brazil and Chile**. Estudios económicos, Universidad de la rioja, ano 2019, v. 36, n. 72, p. 5-30, 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=7389442>. Acesso em: 17 jun. 2020.
- DEMPSTER, M.A.H.; MADAN, Dilip B.; CONT, Rama. **Commodities**. 1. ed. rev. New York: Taylor & Francis Group, LLC, 2016. 703 p. v. 1. ISBN 978-1-4987-1233-0.
- FANG, Yongmei *et al.* Optimal forecast combination based on ensemble empirical mode decomposition for agricultural commodity futures prices. **Journal of Forecasting**, Wiley online library, p. 1, 18 jan. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2665>. Acesso em: 7 abr. 2020.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1. ed. Massachusetts Institute of Technology: The MIT Press, 2016. 775 p. ISBN 9780262035613.

- KIM, Myunghyun. How the financial market can dampen the effects of commodity price shocks. **European Economic Review**, Elsevier B.V., v. 121, p. 1, 11 nov. 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0014292119302004?via%3Dihub>. Acesso em: 7 abr. 2020.
- KINGMA, Diederik P.; BA, Jimmy. Adam: A Method for Stochastic Optimization. **Machine Learning**, Cornell University, 22 jul. 2014. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em: 14 ago. 2019.
- KENT, Harold Baker; FILBECK, Greg; HARRIS, Jeffrey H. **Commodities: Markets, Performance, and Strategies**. 1. ed. New York: Oxford University Press, 2020. 634 p. v. 1. ISBN 0190656018. Disponível em: <https://global.oup.com/academic/product/commodities9780190656010?cc=br&lang=en&>. Acesso em: 1 jun. 2020.
- MARTINS, Carlos. **Os Supersinais da Análise Técnica**: Guia para investimentos lucrativos na bolsa. 1. ed. rev. Rio de Janeiro: Elsevier Editora Ltda, 2010. 133 p. v. 1. ISBN 978-85-352-1755-1.
- MITCHELL, Tom M. **Machine Learning**. 1. ed. New York: McGraw-Hill, 1997. 432 p. v. 1. ISBN 0070428077.
- RIGATOS, Gerasimos *et al.* Forecasting of commodities prices using a multi-factor PDE model and Kalman filtering. **IET Cyber-Physical Systems: Theory & Applications**, IEEE Digital Library, ano 12 2018, v. 3, n. 4, p. 232 - 245, 17 dez. 2018. Disponível em: <https://ieeexplore.ieee.org/document/8577964>. Acesso em: 8 abr. 2020.
- SILVA, Gerarda Beatriz Pinto da. **Comercialização de commodities agrícolas: principais mecanismos**. Instituto Agro, [s. l.], 26 mar. 2019. Disponível em: <https://institutoagro.com.br/commodities-agricolas/>. Acesso em: 7 dez. 2020.
- SRIVASTAVA, Nitish *et al.* Dropout: a simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, Toronto, v. 15, ed. 1, p. 1929-1958, 14 jun. 2014. Disponível em: <https://dl.acm.org/doi/abs/10.5555/2627435.2670313#sec-ref>. Acesso em: 4 dez. 2020.
- TICTACTEC LLC. (Chelmsford, Massachusetts) *et al.* **TA-Lib : Technical Analysis Library**. 1. [S. l.]: TicTacTec LLC. 2007. Disponível em: <https://ta-lib.org/>. Acesso em: 5 dez. 2020.
- UNIVERSIDADE DE SÃO PAULO. **CEPEA - Centro de Estudos Avançados em Economia Aplicada**. Centro de Estudos Avançados em Economia Aplicada. [S. l.], 2020. Disponível em: <https://www.cepea.esalq.usp.br/br>. Acesso em: 20 abr. 2020.
- WANG, Yudong; LIU, Li; WU, Chongfeng. Forecasting commodity prices out-of-sample: Can technical indicators help?. **International Journal of Forecasting**, Elsevier B.V., v. 36, n. 2, p. 666-683, 29 nov. 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169207019302286?via%3Dihub>. Acesso em: 9 abr. 2020.
- YAHOO! - VERIZON MEDIA. **Yahoo Finance**. [S. l.], 2020. Disponível em: <https://finance.yahoo.com/>. Acesso em: 20 abr. 2020.
- YIN, Libo; YANG, Qingyuan. Predicting the oil prices: Do technical indicators help?. **Energy Economics**, Elsevier B.V., v. 56, p. 338-350, 31 mar. 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S014098831630055X>. Acesso em: 7 abr. 2020.
- ZHANG, Dabin *et al.* Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons. **IEEE Access**, IEEE Digital Library, v. 8, p. 28197 - 28209, 4 fev. 2020. Disponível em: <https://ieeexplore.ieee.org/document/8981960>. Acesso em: 9 abr. 2020.

APÊNDICE A – CORRELAÇÃO DE PEARSON DOS INDICADORES TÉCNICOS

A Tabela 7 apresenta os coeficientes de correlação sobre os dados internacionais de mercado. Já a Tabela 8 por sua vez, apresenta os coeficientes de correlação de Pearson sobre os dados do mercado nacional de commodities.

Tabela 7 – Coeficientes de correlação de Pearson sobre os dados internacionais de mercado.

Indicador / Commodity	Açúcar	Algodão	Milho	Soja	Trigo
LEADSINE	- 0,10107002	- 0,11139340	- 0,10212593	- 0,13551039	- 0,06516872
AROON	- 0,10033008	- 0,11097280	- 0,10282447	- 0,11350792	- 0,08271201
HT DCPERIOD	- 0,08391370	- 0,05355143	0,08237644	0,09950873	0,09504813
MINUS_DI	- 0,07458650	- 0,10275092	- 0,36694924	- 0,40552287	0,15746405
Volume	- 0,05004641	0,04439145	0,16360178	0,18584743	- 0,21368455
SINE	- 0,03069817	- 0,00788286	- 0,01367827	- 0,05107745	0,01327946
ADX	- 0,01616075	0,08874568	0,04179882	0,15680072	0,14559721
DX	- 0,01531925	0,03975199	0,01193639	0,08014460	0,06894388
ADXR	- 0,00751500	0,11375709	0,05870809	0,18183309	0,16937756
QUADRATURE	0,00004440	0,00794148	0,00232166	- 0,00012357	0,00177310
HT TRENDMODE	0,01485683	0,01083987	0,02185006	0,12647476	0,08737331
BOP	0,02043372	0,00982113	0,02405297	0,04502379	- 0,00421634
STOCHRSI	0,03208399	0,01335561	0,03358390	0,03447477	0,01257005
INPHASE	0,06561461	0,10133719	0,06305168	0,05345451	0,08872855
STOCHF	0,06680481	0,05560646	0,06996687	0,09330302	0,04319545
WILLR	0,06794297	0,08974563	0,10908313	0,14007553	0,07970797
HT DCPHASE	0,07126735	0,10657388	0,09048341	0,10209420	0,04834393
CCI	0,07740901	0,10643201	0,09318833	0,10159572	0,09272115
STOCH	0,07825254	0,06789033	0,08353473	0,11127261	0,05439451
PLUS DI	0,08129758	0,19734676	- 0,28376611	- 0,31340512	0,38142452
ROC	0,09187217	0,11524043	0,07341352	0,08179887	0,10792474
ROCR	0,09187217	0,11524043	0,07341352	0,08179887	0,10792474
ROCP	0,09187217	0,11524043	0,07341352	0,08179887	0,10792474
ROCR100	0,09187217	0,11524043	0,07341352	0,08179887	0,10792474
ULTOSC	0,09889119	0,07524174	0,09160811	0,18551238	0,08359765
MOM	0,10095672	0,11273025	0,08634693	0,08839602	0,11057594
ADOSC	0,12136935	0,04874322	0,04784867	0,11511327	0,06914568
MIF	0,12366990	0,09393667	0,11558424	0,13772161	0,06408387
RSI	0,12679298	0,15100606	0,11270594	0,13866424	0,13509597
CMO	0,12679298	0,15100606	0,11270594	0,13866424	0,13509597
PPO	0,13041865	0,13623697	0,10279511	0,11139666	0,14287321
APO	0,14986757	0,15358721	0,12571652	0,12625878	0,15835161
MACDEXT	0,14986757	0,15358721	0,12571652	0,12625878	0,15835161
MACD	0,21345477	0,22213166	0,16886301	0,17487327	0,21532759
MACDFIX	0,21350539	0,22209776	0,16885309	0,17488443	0,21533593
AD	0,22007276	- 0,10034715	- 0,01818481	0,19618358	0,11247460
TRIX	0,30061497	0,31872645	0,20487463	0,22235242	0,28089739
NATR	0,31008719	0,35868731	0,31442310	0,28694566	0,11916946
OBV	0,45523242	0,16162160	0,23452564	0,30119409	0,59871290
MINUS DM	0,58859266	0,63013850	0,50869227	0,38265851	0,59956594
TRANGE	0,62805229	0,69649601	0,56482108	0,47438842	0,47674641
PLUS DM	0,73745874	0,83559670	0,67304095	0,61342416	0,75163947
ATR	0,82332262	0,85325667	0,77514359	0,65861361	0,66398429
Open	0,99765305	0,99839716	0,99837778	0,99810809	0,99860156
High	0,99888362	0,99907043	0,99920019	0,99909163	0,99905283
Low	0,99908872	0,99925203	0,99931181	0,99914983	0,99919047
Close	1,0	1,0	1,0	1,0	1,0
Adj Close	1,0	1,0	1,0	1,0	1,0

Fonte: elaborado pelo autor.

Tabela 8 – Coeficiente de correlação de Pearson sobre os dados nacionais de mercado.

Indicador / Commodity	Açúcar	Algodão	Milho	Soja	Trigo
LEADSINE	- 0,23583867	- 0,06569711	- 0,20690492	- 0,23812875	- 0,13165257
SINE	- 0,16595809	0,01027603	- 0,04004380	- 0,10890925	- 0,10089631
STOCHRSI	- 0,12767920	- 0,08985186	- 0,06715591	- 0,01595749	- 0,05882380
QUADRATURE	- 0,07396193	- 0,07253528	- 0,03916635	- 0,00915317	- 0,04586356
HT TRENDMODE	0,03901865	0,10081483	0,13085448	0,06501193	0,04646822
ROCR	0,07177159	0,05981444	0,14999843	0,15932397	0,12367921
ROC	0,07177159	0,05981444	0,14999843	0,15932397	0,12367921
ROCP	0,07177159	0,05981444	0,14999843	0,15932397	0,12367921
ROCR100	0,07177159	0,05981444	0,14999843	0,15932397	0,12367921
MOM	0,08307183	0,06672523	0,16216663	0,16284140	0,11961018
INPHASE	0,09082933	0,09580076	0,17373159	0,14727272	0,16681564
HT DCPHASE	0,15474420	0,05206589	0,19874840	0,20898608	0,11236417
CMO	0,18509653	0,08100763	0,30984300	0,30517003	0,16748187
RSI	0,18509653	0,08100763	0,30984300	0,30517003	0,16748187
PPO	0,18892830	0,15555543	0,24862871	0,27437607	0,22809655
APO	0,19549568	0,16702112	0,26767576	0,28277952	0,23069744
MACDEXT	0,19549568	0,16702112	0,26767576	0,28277952	0,23069744
MACD	0,26502001	0,21018443	0,34247140	0,38870515	0,31003924
MACDFIX	0,26530357	0,21047608	0,34258186	0,38902278	0,31027552
HT DCPERIOD	0,27262300	0,10624508	0,27997955	0,25604055	- 0,02854838
TRIX	0,43366235	0,21416214	0,46843084	0,54283840	0,47789509
Preço	1.0	1.0	1.0	1.0	1.0

Fonte: elaborado pelo autor.