

UNIVERSIDADE FEDERAL DE UBERLÂNDIA - UFU
FACULDADE DE CIÊNCIAS CONTÁBEIS - FACIC
GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS

MATEUS FERREIRA SILVA

**Previsão de Preços no Mercado de Ações Brasileiro por Meio de Técnicas de
Inteligência Artificial**

UBERLANDIA
FEVEREIRO 2022

MATEUS FERREIRA SILVA

Previsão de Preços no Mercado de Ações Brasileiro por Meio de Técnicas de Inteligência Artificial

Projeto de pesquisa apresentado como requisito para obtenção do título de bacharel em Ciências Contábeis pela Faculdade de Ciências Contábeis da Universidade Federal de Uberlândia.

Orientador: Dr. Flávio Luiz de Moraes Barboza

UBERLÂNDIA
FEVEREIRO 2022

RESUMO

Somente em janeiro de 2021, o mercado financeiro brasileiro movimentou mais de R\$35 bi por dia. Isso se deve potencialmente a um aumento recente de participantes no mercado com perfil de operador *intraday* (compras e vendas em um mesmo dia). Porém, apesar de ser um mercado atrativo, há dificuldades na obtenção de lucros consistentes. Em paralelo, houve um aumento de estudos que utilizam análise técnica e fundamentalista combinadas com o aprendizado de máquina para a previsão de tendências futuras nos preços, sendo as redes neurais uma das técnicas mais utilizadas. Esse estudo analisa a qualidade de previsões 15 minutos a frente na tendência (subir ou cair) do preço da ação da Petrobras entre 2014 e 2020. Como existe mais de 100 indicadores técnicos, implementou-se um extrator de variáveis baseado na técnica *eXtreme Gradient Boosting* (XGBoost) para diminuir ruído na amostra. Uma rede neural do tipo *Long Short Term Memory* (LSTM) foi aplicada na etapa de previsão e, como resultado, atingiu-se uma acurácia de 56%. Destaca-se que os melhores resultados ocorreram na previsão de quedas (57%) do que de altas (51%). Em especial, constatou-se que houve melhorias em relação a resultados apresentados anteriormente na literatura.

Palavras-chave: Tendência de Preço, Investimentos, Inteligência Artificial.

ABSTRACT

In January 2021, the Brazilian financial market moved R\$35 billions in a single day. That's might be happened because a recente raise of participants in the Market with an intradayoperator profile (buy and sell assets in the same day). However, even being attractive, there are some issues to obtain consistent profits. At the same time, had a raise in the studies that uses technical analysis and fundamentalism with machine learning to forecast price trends, and the neural network technique more use. This studies analysis the quality to predict 15 minutes forward price trend, using Petrobrás stocks between 2014 and 2020. With More than 100 technical indicators, implemented a variable extractor base on eXtreme Gradient Boosting (XGboost) to reduce noise in the sample. A neural network like Long Short Term Memory (LSTM) was applied to predict the trend price, as a result met 56% accuracy. Emphasisthat the best results occurred to predict downward trend (57%) and upward trend (51%). In particular, it was found that there were improvements in relation to results previously presented in the literature.

Keywords: Price Trend, Investments, Artificial Intelligence

1. INTRODUÇÃO

As operações de *High Frequency Tradings*, *High Frequency Tradings* (HFTs) são assim chamados por realizar operações com alta frequência (minutos ou mesmo segundos), ou seja, realizam um grande número de operações. Essa forma de transacionar causou mudanças substanciais com forte impacto no mercado financeiro nos últimos anos (Malceniace 2019).

Isso se justifica pelo número de pessoas que tem procurado atuar como *day trader* tem aumentado cada vez mais nos últimos anos, na tentativa de ter uma fonte de renda com esse tipo de atividade. Embora isso tenha acontecido, de acordo com Chague et al.(2020), menos de 10% das pessoas conseguem persistir por mais de 300 pregões. Dos que persistem, mais de 97% perdem dinheiro e, surpreendentemente, apenas 1,1% conseguem ganhar mais que 1 salário mínimo do Brasil. O pesquisador esclarece no estudo que esses dados são de investidores na bolsa Brasileira. A partir desse estudo, pode-se observar dificuldades em obter resultados consistentes nesse tipo de operação.

Enquanto isso, a inteligência artificial vem evoluindo e sendo aplicada de diversas formas, por exemplo, para construir banco de dados, reconhecimento de imagens, e, notadamente, para fazer previsões de tendência no mercado financeiro (Nelson et al.2017). Em particular, no mercado financeiro tais técnicas podem ser úteis na previsão de comportamento de ativos, contribuindo para a tomada de decisões de investidores, gestores de fundos, analistas, entre outros. Como maneira alternativa, a máquina pode aprender diretamente com os dados *Machine Learning*, já que essas técnicas de análise podem agilizar (fornecem resultados muitas vezes em milésimos de segundos), e, até mesmo com uma maior precisão e menor custo (Ding et al.2020). Por isso, a utilização dessas tem sido amplamente discutida e melhor analisadas para aplicações em Finanças (Nelson et al. 2017), (Ding e Qing. 2020).

Uma forma considerada mais sofisticada de inteligência artificial é a chamada *Deep Learning*. Essa metodologia pode ser definida como um subcampo de *Machine Learning*, na qual os cientistas se basearam no cérebro humano (redes neurais) para desenvolver métodos pelos quais as máquinas conseguem aprender por si mesmas para superar algumas dificuldades enfrentadas em algoritmos menos complexos (Wall 2018).

No campo da gestão de investimentos, a inteligência artificial vem sendo utilizada por empresas e investidores por conseguir uma análise mais precisa de um alto volume de dados, algo que o mercado financeiro oferece publicamente. Isso envolve desde técnicas estatísticas até aquelas do tipo *ensemble*, conseguindo assim benefícios importantes para seus usuários (Tadapaneni 2019). Em particular, Nelson et al. (2017) relata que 85% das operações no mercado norte-americano advinham de algoritmos.

No mercado financeiro, os preços das ações se movimentam de acordo com as informações, porém elas são, em tese, imprevisíveis. Por conseguinte, os valores dos ativos também, o que leva a pressupor que a oscilação dos preços pode ser considerada aleatória (Rehman et al.2018). Por um lado, *Random Walk* (Passeio Aleatório) é uma visão dada em espaços bidimensionais ou tridimensionais levando em conta conceitos probabilísticos para verificar os possíveis deslocamentos futuros e é geralmente adotada como a melhor previsão possível Henrique et al.(2018). Por outro lado, LSTM é uma rede neural recorrente, onde conseguimos utilizá-la para tempos longos ou curtos. Uma LSTM contém os “portões” de entrada, saída e esquecimento, onde são introduzidas variáveis, realizado um refinamento e por fim obtemos a saída dos dados, durante esse processo, as camadas dessa rede pode acrescentar ou retirar valores de nossas variáveis de acordo com a importância da mesma Selvin et al.(2017).

Pesquisas recentes, tal como Shu (2020) e Ding e Qing (2020), trazem resultados interessantes no emprego de LSTM para previsões de preços e tendências, na tentativa de diminuir a incerteza diante dos dados e fatos. Em particular, Dey (2016) utilizou XGBoost para selecionar variáveis e criar um modelo eficiente para a previsão de ações, chegando a uma

acurácia de 87% em previsões para 60 e 90 dias a frente. Já Nelson et al.(2017) realizou seu estudo em 5 ativos da bolsa de ações brasileira, com previsões do *intraday* de 15 minutos, após coletar os dados de 2014, utilizando uma LSTM se obteve como resultado da acurácia 55,9%, sendo este considerado um bom resultado pelos autores.

Em estudos recentes, pode-se notar que a análise técnica possui tem demonstrado vantagens sobre a análise fundamentalista e análise mista -- uma combinação das duas Henrique et al.(2018), Dai et al(2021). Além disso, redes neurais é um dos algoritmos mais utilizados em *machine learning* para a previsão do mercado de ações (Nti et al.2020).

Diante do exposto, esse estudo aplica a técnica de LSTM para prever a tendência do preço da ação preferencial da Petrobrás (PETR4) com os valores do período de 2014 a 2020, coletando os dados de 15 minutos *day-trading* a partir do Metatrader 5. Os dados foram integrados a uma biblioteca da linguagem *Phyton* (TA-Lib) -- igualmente a Nelson et al. (2017) -- para o processamento das variáveis com base na análise técnica e depois selecionados por meio do XGBoost, seguindo Dey et al. (2016) nessa etapa. Em seguida, a técnica de LSTM fez as previsões de tendência, adotando o mesmo procedimento de (Nelson et al.2017). E por fim, comparamos os resultados com o estudo de (Nelson et al. 2017).

O modelo conseguiu uma melhor performance na previsão de queda do preço do ativo e os resultados revelam que LSTM foi capaz de atingir uma acurácia de 56%, superando o resultado de Nelson et al.(2017).

Essa pesquisa apresenta a contribuição em dois aspectos. Primeiro, mostra avanços em resultados anteriores e segundo, combina técnicas de inteligência artificial para duas etapas distintas: seleção de variáveis e LSTM no mercado brasileiro.

Contudo, percebe-se que essa pesquisa nos traz interessante implicações. Destaca-se a possibilidade de usar tal algoritmo em diversos ativos, como também para diferentes horizontes de tempo, não limitando-se simplesmente a intervalos de 15 minutos.

Este artigo está organizado da seguinte forma. A seção 2 descreve os principais aspectos teóricos que envolve a pesquisa como também alguns trabalhos prévios importantes para sustentar essa pesquisa. Na seção 3 está descrito os dados que foram utilizados, como eles foram coletados e os métodos utilizados para a análise, a seção 4 traz os resultados das análises realizadas e a discussão sobre eles, por fim, a seção 5 mostra as conclusões que foram realizadas com este trabalho.

2. REFERENCIAL TEÓRICO

2.1 Análise de Investimentos de Risco

Para a análise do mercado financeiro, temos duas abordagens principais, sendo elas a análise técnica e a análise fundamentalista (Taylor e Allen 1992). Uma das referências mais conhecidas da análise técnica é Charles Henry Dow, criador da teoria de Dow. Segundo essa teoria, o mercado se move constantemente em tendências. Além disso, Willian Peter Hamilton, sucessor de Dow no *The Wall Street Journal*, afirmava que o movimento que o mercado realiza a todo momento já reflete toda a informação disponível, logo conseguimos fazer uma previsão do futuro observando o passado Brown et al. 1998.

Os analistas técnicos acreditam que o preço e o volume do ativo geram indicadores de movimentações futuras, sendo assim, esses dados são examinados e os retornos podem ser aumentados Blume et al.(1994), eles são parte principal da análise técnica, sendo normalmente plotados nos gráficos com o intuito de prever a tendência futura (Jain et al. 2019). Esses são os indicadores técnicos, os quais podem ser subdivididos nas seguintes categorias (quando observado no software MetaTrader5): Tendência, Volumes e Osciladores. O MetaTrader5 é

uma plataforma digital para investidores, sendo amplamente utilizada atualmente pelo fácil acesso. A Média Móvel e o MACD são os dois indicadores técnicos mais utilizados, ambos utilizam como base médias de valores recentes para seu cálculo, onde o operador consegue escolher o período dessas médias e utilizar combinações entre elas para gerar sinais de comprar ou venda (Khan et al. 2008). Além dos indicadores mais conhecidos, novas formulações são construídas com o objetivo de melhor performance do que os antigos, assim como ocorre em (Dai et al. 2021).

De acordo com Nelson et al (2017), indicadores técnicos são gerados a partir de cálculos matemáticos com o intuito de captar características do ativo com base em seus preços históricos. Nesse estudo, os pesquisadores conseguiram uma acurácia de 55,9% para previsões com intervalos de 15 minutos. Tal resultado pode ser considerado expressivo para operações do tipo *intraday* (compra e venda de ativos dentro do mesmo dia de negociação). Em contrapartida, é necessário um dispêndio de tempo para se investir dessa forma, uma vez que observar o mercado em intervalos razoavelmente curtos -- como é o caso de Nelson et al. (2017) -- requer praticamente dedicação exclusiva como *trader*.

2.2 Inteligência Artificial para Previsões de Tendência no Mercado Financeiro

LSTM é um tipo de rede neural recorrente, sendo uma classe de redes neurais em que os processos computacionais formam um ciclo direcionado, e que possui uma arquitetura com valores em intervalos arbitrados (Selvin et al.2017). Sua aplicação em finanças vem demonstrando eficácia maior que outros métodos, principalmente para prever as tendências em séries temporais de preços. Os resultados de estudos mais recentes demonstram que o uso dela tem trazido melhor desempenho devido a forma como foi estruturada e desenvolvida, armazenando melhor as informações mais antigas a medida que são transmitidas por cada neurônio Nelson et al. (2017). Devido a arquitetura da LSTM, diversas pesquisas mostram que ela é perspicaz em diferentes intervalos temporais, podendo ser utilizada em modelos de curto, médio e longo prazo, e também pode ser utilizada em modelos de classificação e regressão (Selvin et al.2017).

Atualmente existem uma variedade pesquisas que fundamentam a LSTM como um excelente modelo de inteligência artificial, o que traz bastante inovação e resultados cada vez mais promissores na área, como o *particle swarm optimization* que é um parâmetro que auxilia as redes neurais e pode melhorar a performance do modelo ao custo de um alto tempo de processamento, porém está comprovado que o modelo pode melhorar bastante (Ding e Qing 2020). Dessa forma devido alto tempo de processamento torna-se inviável utilizar em previsão para movimentações de alta frequência.

Outro modelo utilizado para a realização do estudo foi o XGBoost, que utiliza um método de classificação baseado nos conceitos de árvore de decisão, porém com algumas modificações que o torna um modelo mais completo com relação a uma árvore de decisão comum Dey (2016), sendo um classificador composto por pequenas árvores de decisão treinando sequencialmente novas árvores e aprendendo com os erros das árvores anteriores, o que torna o modelo um pouco mais eficiente que uma árvore de decisão comum. O modelo também demonstrou resultados interessantes para realizar a extração de variáveis Zhang et al.(2014), sendo um procedimento realizado no pré-processamento dos dados antes de chegar aos resultados finais do estudo.

A Tabela 1 apresenta alguns resultados interessantes na área, assim como algumas peculiaridades de cada estudo.

Tabela 1. Estudos que empregaram inteligência artificial para fazer previsões de preços.

Estudo	Objetivo	Extração de Variáveis	Variáveis de Análise Técnica	Técnicas usadas
--------	----------	-----------------------	------------------------------	-----------------

Dey (2016)	Prever Tendência	✓	RSI, STOCH, OBV, MACD	XGBoost
Nelson et al.(2017)	Prever Tendência	✗	Todas	LSTM, Random Forest
Henrique et al.(2018)	Prever o Valor	✗	SMA, WMA, RSI, ADO, ATR	SVR
Grigoryan et al.(2015)	Prever o Valor	✓	EMA, OBV, ADX, DX, MACD	NARX, PCA, RNN
Karathanasopoulos et al.(2016)	Rendimento do modelo	✓	MACD	ARMA, MLP, ESVM

2.3 Random Walk

O modelo *Random Walk* propõe que o preço futuro das ações não tem relação alguma com seu histórico passado Fama (1965) e que o mercado financeiro age de maneira totalmente aleatória, dessa forma não faria sentido tentar prever quais os valores futuros usando como base os dados anteriores. Diferentemente da hipótese do mercado eficiente, onde diz que o mercado segue tendências e se ajustam de formas que os ganhos entre os investidores sejam padronizados e que para aumentar os ganhos o investidor deve correr mais riscos (Fama e Malkiel 1970).

Pode-se então definir como parâmetro que a média do mercado é a melhor hipótese de ganho e qualquer modelo que ultrapasse essa média pode ser considerado um modelo eficiente. Dessa forma foi desenvolvida a equação em que a melhor forma de prever o próximo intervalo temporal é colocar como previsão o que aconteceu no período anterior, dessa forma esses resultados costumam ser utilizados como base de comparação para modelos de classificação em inteligência artificial onde utilizando os mesmos dados concluir se o modelo é satisfatório ou não (Henrique et al. 2018).

3. METODOLOGIA

Esse estudo realiza uma análise quantitativa e exploratória, uma vez que estudos exploratórios podem ser realizados com a utilização de fontes primárias buscando seus dados históricos, esses estudos podem ainda ser caracterizados tanto como quantitativos quanto qualitativos, estudos quantitativos utilizam diferentes métodos estatísticos para quantificar os dados obtidos (Marconi e Lakatos, 1990).

3.1 Dados

Para a realização do estudo, utilizamos dados da Petrobras -- em particular, preços da ação preferencial, código PETR4 -- companhia do setor de petróleo e gás, uma das principais ações componentes do Índice Bovespa. Os dados foram obtidos por meio da plataforma Metatrader 5, pela qual é possível automatizar a extração de seus dados integrando a uma biblioteca utilizada no processamento em linguagem *Python*.

Os valores coletados de preços (abertura, fechamento, mínimo e máximo) e volume de negócios foram extraídos no ano de 2014 (ou melhor, de 01/01/2014 a 31/12/2014), sendo estes dados com um intervalo de tempo (conhecido como *timeframe*) de 15 minutos, ou seja, a cada 15 minutos apresenta-se cinco dados a serem coletados. Dessa forma, foram extraídas 6848 observações, sendo todas elas sem nenhum dado faltante. A Tabela 2 apresenta uma amostra com as 5 primeiras e com as 5 últimas amostras dos dados coletados. Esta base [extraída especificamente em 2014 e usando intervalos de 15 minutos] é justificada a fim de permitir a comparação direta com o trabalho Nelson et al.(2017), que realizaram seu estudo com este mesmos dados. Adicionalmente, a análise é replicada para os anos de 2015 a 2020, a fim de

verificar se evitar o viés de seleção em torno do intervalo temporal dos dados.

Tabela 2. Dados Brutos extraídos dos valores da ação da Petrobrás no Metatrader 5. O período está expresso no formato ``DD-MM-YYYY HH:MM:SS''. O formato de descrição segue o padrão adotado pela biblioteca de extração de dados. Com isso, a ideia é facilitar a replicação e visualização por parte do leitor.

Período	Abertura	Máximo	Mínimo	Fechamento	Volume
03-01-2014 10:10:00	14.74	14.74	14.69	14.74	24900
03-01-2014 10:15:00	14.74	14.78	14.73	14.73	488900
03-01-2014 10:30:00	14.73	14.75	14.68	14.70	434600
03-01-2014 10:45:00	14.70	14.70	14.63	14.64	339600
03-01-2014 11:00:00	14.64	14.64	14.56	14.61	61600
...					
30-12-2014 15:45:00	9.53	9.63	9.53	9.58	1085400
30-12-2014 16:00:00	9.59	9.59	9.51	9.52	752900
30-12-2014 16:15:00	9.52	9.53	9.46	9.49	1133700
30-12-2014 16:30:00	9.49	9.52	9.41	9.43	1795700
30-12-2014 16:45:00	9.43	9.46	9.39	9.39	6150500

A variável de interesse (Y) foi definida em termos de tendência e, por isso, é vista por apenas dois valores (0 ou 1), considerada então uma variável binária. Assim, Y recebe o valor 1 caso o preço de fechamento do período $t+1$ for maior que o valor no período t . Caso contrário, Y vale 0. Essa mesma notação é usual em trabalhos de previsão de tendência (Henrique et al. 2018; Nelson et al. 2017, Dey 2016).

3.2 Indicadores Técnicos e a Seleção de Variáveis

A partir dos dados extraídos do MetaTrader 5, elaborou-se um conjunto de dados com um total de 113 variáveis baseadas em indicadores de análise técnica fornecidas pela biblioteca TA-Lib Nelson et al. (2017) que abrangem diversas categorias, tais como, estudos de sobreposição, indicadores de momento, indicadores de volatilidade, transformador de preço, indicador de ciclos, reconhecimento de padrões, funções estatísticas, operadores matemáticos e transformadores matemáticos, além das métricas mais empregadas, como é o caso de média móvel convergente e divergente (conhecido como MACD), estocástico, bandas de Bollinger.

Para o desenvolvimento de modelos de previsão em inteligência artificial utilizou-se uma grande quantidade de variáveis, onde muitas dessas são irrelevantes para os cálculos realizados pelo modelo e tendem a piorar os resultados obtidos. Por exemplo, Zhang et. (2014) examinou o processo de extração de variáveis para a previsão de tendência preços de ativos na Bolsa chinesa de Shanghai.

Para isso utilizou-se uma outra técnica de *machine learning* denominada XGBoost, a mesma empregada no estudo de previsão para classificação binária feito por (Dey 2016). Dessa maneira, ao construir o extrator ele foi capaz de fazer um ranking das variáveis que são mais relevantes, e por meio desse ranking, realizou-se uma seleção das variáveis que melhor contribuíram para explicar a tendência futura do ativo (Y).

O ranqueamento foi realizado de tal forma que se atribui uma porcentagem de relevância

para cada uma das 113 variáveis (obviamente, a soma dos percentuais representa um total de 100 por cento). Com isso, foi feito um filtro para a seleção das variáveis que entram na próxima etapa da pesquisa. Para tanto, efetuou-se um corte em 2%, sendo que apenas indicadores com importância acima deste ponto de corte foram selecionados. O resultado deste filtro foi que 8 variáveis serão utilizadas no modelo de previsão.

A Tabela 3 apresenta as variáveis selecionadas, considerando o maior grau de importância. Outros estudos que também investigaram tais métricas com o mesmo objetivo (ou até mesmo próximo ao que pretende-se nesse estudo) são elencados como referências de cada uma delas.

Tabela 3. Variáveis de Análise Técnica que foram selecionadas a partir do XGBoost: *Pearson -Pearson's Correlation Coefficient; Hilbert Transform – SineWave; Absolute Price Oscillator (APO); Commodity Channel Index(CCI); Directional Movement Index (DMI); Moving Average Convergence/Divergence (MACD); Stochastic(STOCH); Ultimate Oscillator (ULTOSC)*

Variável	Definição	Categoria	Referência
Pearson	Mede o grau de correlação entre duas variáveis de escala métrica	Estatístico	Karathanasopoulos et al. (2016)
HT-SineWave	Modelo matemático definido a partir da transformada de Hilbert	Cíclico	Shu e Gao (2020)
APO	Resultado da diferença entre duas médias móveis em termos absolutos	Momento	Grigoryan et al.(2015)
CCI	É a diferença entre o preço atual e a média do período	Momento	Grigoryan et al.(2015)
DMI	Identifica qual a tendência o ativo está seguindo	Momento	Grigoryan et al.(2015)
MACD	Analisa a convergência e divergência entre duas médias móveis para identificar tendências de preço	Momento	Dey (2016)
STOCH	Compara o preço de fechamento do período com relação ao preço médio de períodos anteriores	Momento	Dey (2016)
ULTOSC	Mede o impulso em três intervalos de tempo variáveis com o intuito de corrigir os outros Indicadores de momento	Momento	Grigoryan et al.(2015)

Interessante observar que desses 8 indicadores, 6 são classificados como indicadores de momento. A partir dessa seleção de variáveis, realizou-se a construção de um modelo que retorna os resultados de maneira (supostamente) mais rápida e eficiente para realizar as operações de alta frequência.

3.3 Modelos Previsores

Após o processo inicial de extração e processamento dos dados, foi dado início ao desenvolvimento do algoritmo do modelo a ser utilizado. O primeiro passo foi o escalonamento e a remodelagem dos dados das 8 variáveis selecionadas, sendo denominada como X, e a tabela de valores binários para realização da nossa classificação foi denominada com Y, separando o X e Y em 75% dos dados para treino 25% dos dados para teste. Cabe ressaltar que a separação foi realizada de forma cronológica (os primeiros dados estão presentes na amostra de treino), ou seja, sem sorteio das amostras. Esse formato de divisão das amostras permite que as previsões sejam utilizadas no mundo real, pois esta organização está baseada em um modelo treinado a partir de dados do passado para realizar previsões com dados cronologicamente a frente daqueles usados para construir o modelo. Em contrapartida, é esperado que haja uma perda na qualidade das previsões, uma vez que algoritmos de previsão são mais assertivos quando os dados são embaralhados temporalmente. Isto faz com que sua aplicação seja suspeita e possa prejudicar o emprego do estudo no dia-a-dia do investidor.

Para construirmos a rede neural utiliza-se a biblioteca Keras, optando pela construção

de apenas uma camada, tendo como principais parâmetros 20 unidades e como função de ativação a tangente hiperbólica, conforme (Nelson et al. 2017). Também foi adicionado um *dropout* de 0.2 com o intuito de regularizar o aprendizado da máquina e consequentemente evitar que o nosso modelo tenha o chamado *overfitting* (Srivastava et al 2014), que é quando o modelo se adapta tão bem aos dados, porém ele acaba se tornando ineficaz em prever novos resultados. Para a compilação do modelo foram utilizados o otimizador adam, como validação de erro *binary crossentropy* já que se trata de um modelo de classificação binária e por último a métrica a ser avaliada é a acurácia, que será utilizado como parâmetro de eficácia do nosso modelo e partir disso o modelo foi treinado por 500 épocas.

3.4 Análise de Desempenho

A qualidade do modelo é dada por meio do *Classification Report* -- disponível na biblioteca *scikit-learn* -- o qual calcula e expõe as métricas utilizadas como base de comparação entre o modelo LSTM e o *Random Walk*. No caso, essas medidas são: *Accuracy* que indica quantas previsões o modelo classificou de forma correta como um todo; *Precision* que dentre todas as previsões de tendência no preço que o modelo fez, quantas estão corretas. No caso do mercado de ações o modelo necessita ter uma boa precisão, uma vez que mal utilizado pode gerar perdas financeiras; *Recall* mede quantas previsões foram acertadas em relação às situações reais; e, *F1-Score* é a média harmônica entre *Precision* e *Recall*, sendo também uma forma de avaliar o modelo no geral, uma vez que o resultado dela é refletido dos resultados das outras duas métricas (Nelson et al. 2017). A Tabela 4 apresenta a memória de cálculo de cada uma delas.

Com relação a consistência do modelo, as curvas de erro nas amostras treino e validação foram analisadas antes de se examinar os resultados em si, como forma de confirmar o devido ajuste do modelo.

Tabela 4. Descrição das fórmulas das métricas utilizadas. Legendas: *tp* = true positive *tn* = true negative *fn* = false negative *tn* = false positive

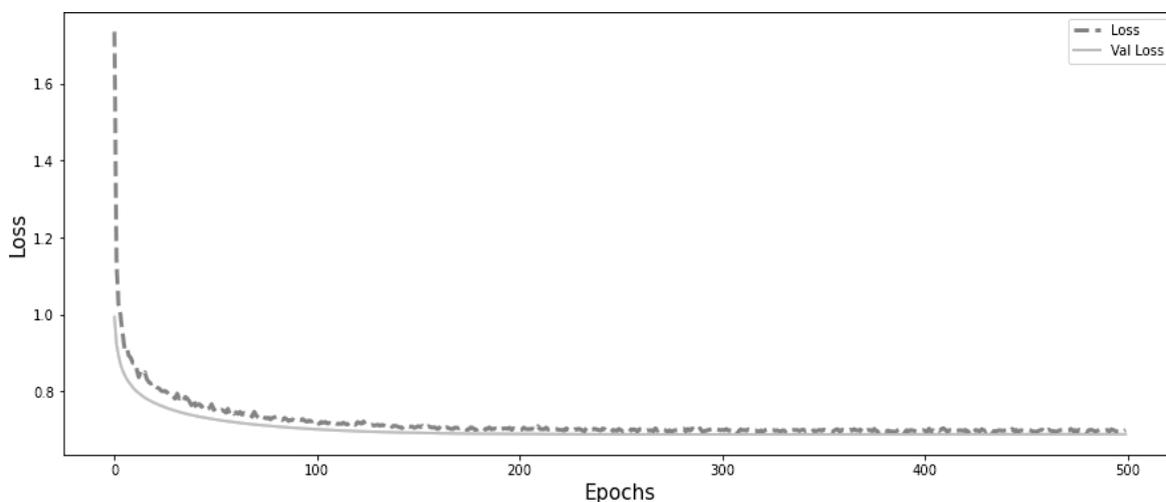
Medida	Fórmula
Acurácia	$A = (tp+tn) / (tp+fp+tn+fn)$
<i>Precision</i>	$P = tp / (tp+fp)$
<i>Recall</i>	$R = tp / (tp + fn)$
<i>F1-Score</i>	$2 * (P * R) / (P + R)$

4. RESULTADOS E DISCUSSÃO

4.1 Consistência do Modelo

Após a realização do treinamento do modelo os primeiros resultados buscados foram as medidas de perdas -- para as duas amostras -- uma vez que elas têm uma relação direta com a qualidade do modelo, ou seja, como está se comportando com relação ao *Overfitting*, que pode ser identificado quando a perda na amostra teste está maior que o no treino. Em contrapartida, pode gerar *Underfitting*, que acontece quando a perda na amostra teste está menor que na outra amostra. De acordo com as curvas de perdas mostradas no gráfico a seguir (Figura 1), o modelo proposto não apresentou nenhum problema com nenhuma das duas medidas.

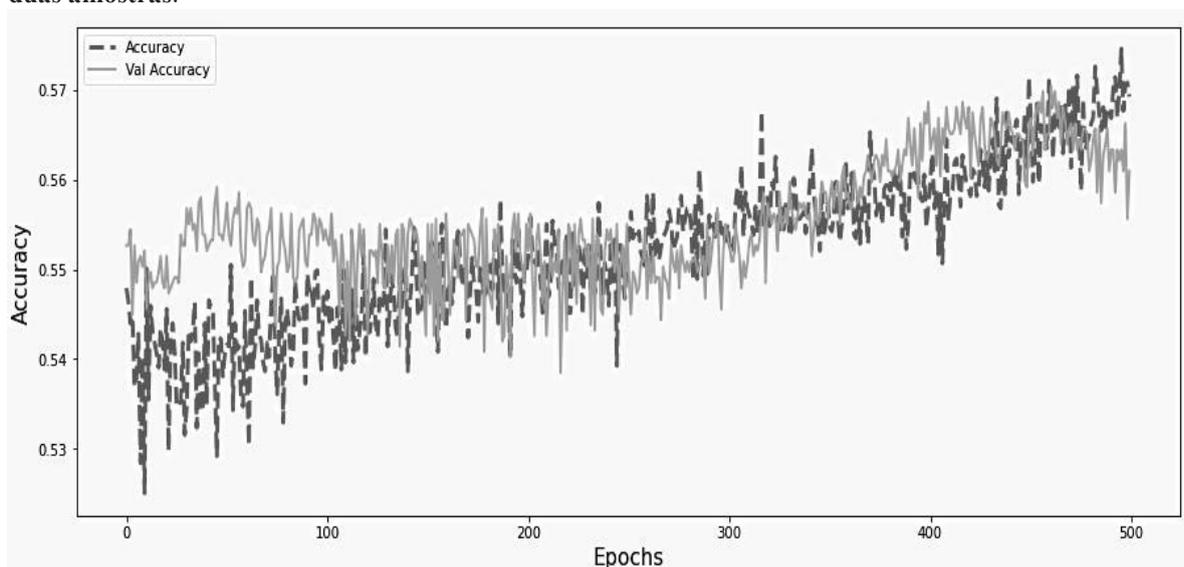
Figura 1 - Gráfico Loss (Perdas) x Epochs (Épocas), evidenciando a consistência do modelo de LSTM.



Com o gráfico percebe-se que o modelo se ajustou aos dados inseridos nele e conseguiu manter o treino do modelo de maneira eficiente sem nenhuma variação considerada substancial nos resultados ao longo das épocas.

A segunda métrica a ser avaliada é a acurácia, sendo uma das mais importantes, uma vez que é ela que demonstra no geral o quanto o modelo está acertando. O gráfico mostrado na Figura 2 (a seguir) descreve a evolução da acurácia do modelo. Ao realizar o treino do modelo, foi possível observar como o algoritmo evoluir em termo de qualidade de previsão. Em suma, o modelo não apresentou melhorias relevantes através das épocas, iniciando com um grau de acurácia de 55%, e chegando a um pico de 57%, o que torna o modelo um pouco melhor do que realizar simplesmente o lançamento de uma moeda. Porém, por ser uma operação com um pequeno intervalo de tempo, essa acurácia pode tornar o modelo viável.

Figura 2 - Gráfico Accuracy (Acurácia) x Epochs (Épocas), mostrando o desenvolvimento do modelo nas duas amostras.



Fonte:

4.2 Classification Report

Para avaliar melhor o modelo, foram utilizadas as métricas encontradas no *Classification Report*, em particular, *Precision*, *Recall*, *F1-score* e Acurácia. No primeiro caso, foi avaliado como os dados se comportam ao utilizar o modelo *Random Walk*, conforme descrito na Tabela 5 temos uma acurácia de 0.51 que seria a mesma coisa que tentar adivinhar. Numa análise mais detalhada, vemos que a precisão mostra que, caso se previsse qual seria a próxima tendência a partir da tendência atual, a técnica de *Random Walk* teria maior assertividade para detectar as baixas do valor de mercado, enquanto as altas do valor de mercado ficaram abaixo de 50%, o que pode ser considerado um desempenho ruim e sem contribuição prática para um investidor (isto é, jogar uma moeda seria mais eficaz).

Já quando avaliamos como os dados se comportam quando é utilizado o modelo de LSTM conforme descrito na Tabela 5 vemos uma melhora de 0.05 na acurácia, chegando a 0.56 e quando olhamos a precisão dele para prever quedas temos 0.57, porém também temos um *recall* de 0.79 e dessa forma podemos perceber que o modelo é mais eficaz para prever as quedas do valor de mercado do ativo.

Tabela 5. Relatório de Classificação para a técnica de LSTM e Random Walk para prever a tendência do preço da ação da Petrobrás na amostra teste. A última coluna (Suporte) mostra a quantidade de vezes que ocorreram os eventos de queda e alta.

Tendência de Preço	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Suporte
LSTM				
Queda(0)	0.57	0.79	0.66	934
Alta(1)	0.51	0.27	0.35	756
Acurácia Total	56%			1690
<i>Random Walk</i>				
Queda(0)	0.55			933
Alta(1)	0.45			756
Acurácia Total	51%			1689

A tabela 6 mostra os resultados obtidos pelo Nelson et al. (2017) onde ele também desenvolveu um modelo LSTM e analisou dados de 15 minutos de algumas empresas presentes na bolsa de valores brasileira, em seu relatório de classificação podemos estabelecer uma comparação direta com a acurácia do modelo, em que ele obteve um total de 53.3%, sendo apenas 3% menor que a performance do modelo proposto.

Já quando comparado com a precisão que foi de 0.563 Nelson et al. (2017) obteve um resultado um pouco melhor na média entre altas e quedas, o que torna o modelo dele um pouco melhor nesse quesito quando leva-se em consideração o modelo como um todo. Porém já no *Precision* e no *Recall* o modelo apresentou resultados superiores quando se comparado ao modelo de (Nelson et al. 2017).

Analisando os dois modelos, percebe-se que ambos que não tiveram uma acurácia total muito acima dos 50%, o que pode indicar uma limitação dos modelos a se adaptar aos dados com um intervalo temporal pequeno. Também foram analisados os anos subsequentes do modelo proposto no estudo como demonstrado na tabela 6 e a partir dele percebe-se que os resultados são bem constantes ao longo dos anos, como nas medidas de *Precision* e *Recall* em que mantiveram uma média de 52.1 e 55.1 respectivamente e que representa a capacidade de classificar corretamente os dados que foram inseridos no modelo. A medida de acurácia apresentou uma certa constância também nos resultados, apresentando seu melhor resultado no ano de 2015 com 59.8% e seu pior desempenho no ano de 2020 com 53.1% e uma média geral de 55.24% o que ainda traz um resultado melhor quando se comparada com Nelson et al.(2017) para o ano de 2014.

Tabela 6. Comparativo entre o modelo proposto -- que usa a técnica de LSTM após selecionar variáveis com XGBoost -- com desempenho nos anos de 2014 até 2020 para o ativo PETR4 e o resultado de Nelson et al.(2017) -- que analisou somente no primeiro ano e com o mesmo ativo.

Modelo de Previsão	Ano	Acurácia	Precision	Recall	F1-score
Nelson et al.(2017)	2014	53,30%	0.56	0.23	0.32
Modelo Proposto	2014	56,00%	0.543	0.557	0.551
	2015	59,80%	0.528	0.599	0.561
	2016	55,9%	0.533	0.554	0.543
	2017	54,4%	0.521	0.564	0.542
	2018	54,3%	0.513	0.539	0.526
	2019	53,2%	0.503	0.526	0.514
	2020	53,1%	0.512	0.523	0.518
	Média	55,24%	0.521	0.551	0.536

O modelo apresentou um melhor desempenho para prever as quedas do mercado financeiro em todos os anos e vale ressaltar também que ao realizar todo o pré-processamento dos dados em todos os anos, o extrator de variáveis selecionou sempre as mesmas 8 variáveis de análise técnica como as mais importantes, demonstrando também a consistência XGBoost ao realizar o ranking das variáveis mesmo quando utilizamos períodos distintos.

5. Conclusões

Diante de um mercado que movimenta milhões de reais em um único dia, este estudo examina a capacidade preditiva de um modelo de inteligência artificial (no caso, LSTM) para prever a tendência 15 minutos a frente de preços das ações da Petrobras nos anos de 2014 a 2020, usando indicadores técnicos selecionados a partir da técnica do XGBoost.

A performance deste algoritmo foi então comparada com a técnica de *Random Walk*, onde é possível perceber que uma maior acurácia no modelo proposto (56% ante 51%). Além disso, o resultado evidencia maior eficácia quando se é utilizado para prever as tendências de queda 57% sendo que foi capaz de acertar 79% destes eventos, o que dá ampla possibilidade de aplicação prática para o modelo proposto. Cabe ressaltar que, na comparação com estudo de Nelson et al.(2017) este resultado trouxe melhorias. Isso mostra que a seleção de variáveis trouxe benefícios para a melhor performance do preditor.

Nesse sentido, o trabalho apresenta duas principais contribuições. Uma delas no âmbito científico, onde mostra o aumento de qualidade de modelos previsores de tendência no mercado financeiro por meio de seleção de atributos, o que também diminui custo computacional. A segunda contribuição refere-se a potencial aplicabilidade deste previsor em produtos financeiros, abrindo possibilidade de uso em uma vasta carteira de ativos, assim como na gestão de riscos de mercado.

O trabalho apresenta dois fatores limitantes em seu desenvolvimento, sendo o primeiro deles a aplicação do modelo em apenas um único ativo. Outra limitação foi o uso de apenas indicadores técnicos a partir da biblioteca TA-Lib, apesar desta biblioteca contemplar grande parte desses indicadores. Porém, o estudo pode

futuramente abarcar variáveis baseadas em indicadores de governança, fundamentos e até mesmo notícias.

Os próximos passos a se desenvolver no estudo seria a utilização do modelo para realizar simulações, com o intuito de colocá-lo como tomador de decisões em uma carteira de investimentos e analisar os resultados com base em medidas específicas da gestão de investimentos, tais como, o beta e outros indicadores de risco.

6. Referências

BLUME, L.; EASLEY, D.; O'HARA, M. Market statistics and technical analysis: The role of volume. **The Journal of Finance**, Wiley Online Library, v. 49, n. 1, p. 153–181, 1994.

BROWN, S. J.; GOETZMANN, W. N.; KUMAR, A. The dow theory: William peterhamilton's track record reconsidered. **The Journal of Finance**, Wiley Online Library, v. 53, n. 4, p. 1311–1333, 1998.

CHAGUE, F.; GIOVANNETTI, B. É possível viver de day-trade em ações? **Brazilian Review of Finance**, v. 18, n. 3, p. 1–4, 2020.

DAI, Z.; ZHU, H.; KANG, J. New technical indicators and stock returns predictability. **International Review of Economics and Finance**, v. 71, p. 127–142, 2021.

DEY, S. Forecasting to classification: Predicting the direction of stock market price using xtreme gradient boosting. **Working Paper**, 2016.

DING, G.; QIN, L. Study on the prediction of stock price based on the associated network model of lstm. **International Journal of Machine Learning and Cybernetics**, v. 11, p. 1307–1317, 2020.

FAMA, E. F. Random walks in stock market prices. **Financial Analysts Journal**, **Routledge**, v. 21, n. 5, p. 55–59, 1965.

GRIGORYAN, H. et al. Stock market prediction using artificial neural networks. case study of tallt, nasdaq omx baltic stock. **Database Systems Journal**, Academy of Economic Studies-Bucharest, Romania, v. 6, n. 2, p. 14–23, 2015.

HENRIQUE, B. M.; SOBREIRO, V. A.; KIMURA, H. Stock price prediction using support vector regression on daily and up to the minute prices. **The Journal of Finance and Data Science**, v. 4, n. 3, p. 183–201, 2018.

JAIN, V.; SAINI, D.; AHLUWALIA, A. Real-time autonomous trading system. **Journal of Statistics and Management Systems**, Taylor Francis, v. 22, n. 2, p. 403–413, 2019.

KARATHANASOPOULOS, A. et al. Stock market prediction using evolutionary support vector machines: an application to the ase20 index. **The European Journal of Finance**, Routledge, v. 22, n. 12, p. 1145–1163, 2016.

Khan, A. U.; Bandopadhyaya, T. K.; Sharma, S. Comparisons of stock rates prediction accuracy using different technical indicators with backpropagation neural network and genetic algorithm based backpropagation neural network. In: **2008 First International Conference on Emerging Trends in Engineering and Technology**, p.575–580, 2008.

MALCENIECE, L.; MALCENIEKS, K.; PUTNINŠ, T. J. High frequency trading and comovement in financial markets. **Journal of Financial Economics**, v. 134, n. 2, p. 381–399, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304405X19300832>>.

MALKIEL, B. G.; FAMA, E. F. Efficient capital markets: A review of theory and empirical work. **The Journal of Finance**, v. 25, n. 2, p. 383–417, 1970.

MARCONI, M. D. A., & LAKATOS, E. M. **Técnicas de pesquisa**. 3ª ed. São Paulo: Atlas, 1990.

Nelson, D. M. Q.; Pereira, A. C. M.; de Oliveira, R. A. Stock market's price movement prediction with lstm neural networks. In: **2017 International Joint Conference on Neural Networks (IJCNN)**, 2017. p. 1419–1426.

NTI, I. K.; ADEKOYA, A. F.; WEYORI, B. A. A systematic review of fundamental and technical analysis of stock market predictions. **Artif Intell Rev** **53**, p. 3007–3057, 2020.

REHMAN, S. et al. Are stock prices a random walk? an empirical evidence of asian stock markets. **ETIKONOMI**, v. 17, n. 2, p. 237–252, 2018.

Selvin, S. et al. Stock price prediction using lstm, rnn and cnn-sliding window model. In: **2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**, 2017. p. 1643–1647.

SHU, W.; GAO, Q. Forecasting stock price based on frequency components by emd and neural networks. **IEEE Access**, v. 8, p. 206388–206395, 2020.

SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

TADAPANENI, N. R. Artificial intelligence in finance and investments. **Biographical Memoirs of Fellows of the Royal Society**, p. 2792–2795, 2019.

TAYLOR, M. P.; ALLEN, H. The use of technical analysis in the foreign exchange market. **Journal of international Money and Finance**, Elsevier, v. 11, n. 3, p. 304–314, 1992.

WALL, L. D. Some financial regulatory implications of artificial intelligence. **Journal of Economics and Business**, v. 100, p. 55–63, 2018. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0148619517302618>>.

ZHANG, X. et al. A causal feature selection algorithm for stock prediction modeling. **Neurocomputing**, v. 142, p. 48–59, 2014.