VETOR

Previsão de Tendência de Preço no Mercado Brasileiro de Ações usando Cadeias de Markov de Tempo Discreto

Price Trend Forecasting in the Brazilian Stock Market using Discrete-Time Markov Chain

João Queiroz[1], Martha Torres[1,†]

[1]*Universidade Estadual de Santa Cruz, Ilhéus, Brasil*
[†]**Autor correspondente:** mxtd@uesc.br

Resumo

O entendimento da tendência do mercado de ações com o objetivo de prever o movimento de preço é muito importante para decisões de investimento dado que os preços das ações são afetados não somente pelo estado financeiro da empresa, mas também por condições políticas, sociais, econômicas, globais e locais, além de muitos outros fatores. As Cadeias de Markov proporcionam um ferramental poderoso para realizar modelagem matemática e computacional e têm sido também usadas para prever tendências no mercado de ações. A partir disto, o seguinte trabalho traz uma ferramenta computacional modelada com base nos conhecimentos obtidos através dos estudos sobre Cadeias de Markov de tempo discreto capaz de realizar previsões de tendências de preço de ações da bolsa de valores brasileira utilizando o método de 3 estados. Foram realizadas análises em 50 ações da BOVESPA a fim de observar se o percentual de sucesso das previsões tem alguma relação com o tamanho do período para construção da matriz de transição. Estes testes foram realizados para os anos de 2019, 2020 e 2021 com o objetivo de observar se houve impactos na efetividade dos métodos durante o período de pandemia de COVID-19.

Palavras-chave

Cadeias de Markov de Tempo Discreto • Previsão de séries temporais • Previsão do mercado de ações

Abstract

Understanding the stock market trend in order to predict price movement is very important for investment decisions given that stock prices are affected not only by the financial state of the company, but also by political, social, economic, global and local, plus many other factors. Markov Chains provide a powerful tool for performing mathematical and computational modeling and also have been used to predict trends in the stock market. From this, the following work brings a computational tool modeled based on the knowledge obtained through studies on discrete-time Markov Chains capable of making predictions of price trends of stocks on the Brazilian stock exchange using the 3-state method. Analyzes were carried out on 50 BOVESPA stocks in order to observe whether the forecast success percentage has any relation to the length of the period for building the transition matrix. These tests were carried out for the years 2019, 2020 and 2021 in order to observe whether there were impacts on the effectiveness of the methods during the period of the COVID-19 pandemic.

Keywords

Discrete-Time Markov chains • Time Series Prediction • Stock Market Prediction

1 Introdução

O mercado de ações mundial movimenta somas gigantescas de dinheiro diariamente. Em maio de 2022, a bolsa de valores do Brasil, B3, movimentou R$ 30,4 bilhões por dia. A previsão de retorno do mercado de ações é uma tarefa desafiadora, dado que os preços das ações são afetados não somente pelo estado financeiro da empresa, mas também por condições políticas, sociais, econômicas globais e locais, além de muitos outros fatores.

Os investidores em geral têm demonstrado grande interesse em ferramentas que ajudam a prever as tendências do mercado de ações. Essas ferramentas buscam prever índices e preços de ações com a maior precisão possível. Atualmente existem diferentes técnicas desenvolvidas para fazer isto, mas devido à volatilidade do mercado, nenhum método em particular pode prever exatamente mudanças no mercado de ações diariamente.

O modelo de Cadeias de Markov de tempo discreto tem sido aplicado para prever tendências de mercado de ações seja através de índices do mercado ou preço de fechamento de ações, utilizando métodos que usam de intervalos de preço de fechamento para construção de estados ou do método de três estados, onde os estados são: o preço aumenta, o preço diminui ou permanece igual e também calculando o comportamento de estado estável.

Tanto os autores que desenvolvem a modelagem com base em índices [1,2] quanto os que desenvolvem com base no preço de fechamento [1,3-6] afirmam que os modelos baseados em Cadeias de Markov auxiliam investidores locais e externos a identificarem tendências futuras de ações e índices, servindo como um excelente indicador para tomada de decisões de investimento.

O presente trabalho traz resultados obtidos através do desenvolvimento de uma ferramenta web que implementa um método baseado em Cadeias de Markov de tempo discreto utilizando 3 estados para construção da matriz de probabilidade de transição e faz previsão de tendência de preço. A análise será feita levando em consideração 50 ações da B3, considerando dados de 2010 até 2021. As questões a serem respondidas são se há alguma relação entre período e precisão das previsões, se é possível melhorar o número de acertos aumentado ou diminuindo os períodos de treinamento, bem como saber se houve alguma mudança na eficácia dos métodos de previsão em períodos atípicos como os últimos anos acometidos pela pandemia de COVID-19, onde o ano de 2019 é considerado pré-pandêmico, o de 2020 pandêmico e o de 2021 pós-pandêmico.

O resto do artigo está organizado como segue. No item 2 será relatado o contexto teórico da pesquisa realizada, enquanto que no item 3 será apresentado o estado da arte deste trabalho. No item 4 será explicada a metodologia aplicada para a obtenção de resultados. No item 5 será descrito o desenvolvimento e no item 6 mostram-se os resultados obtidos. Já no item 7 apresentam-se as conclusões e trabalhos futuros deste trabalho.

2 Contexto Teórico

Aqui serão descritos alguns conceitos fundamentais de cadeias de Markov de tempo discreto, tomando como base o livro de Kulkarni [7].

Considere um sistema que evolui de maneira aleatória no tempo, suponha que o sistema é observado em tempos $n = 0, 1, 2, 3, ...$ Seja Xn o estado do sistema no tempo n, a sequência de variáveis aleatórias $\{X_0, X_1, X_2, ...\}$ é chamado de um processo estocástico de tempo discreto e é escrito como $\{Xn, n \geq 0\}$. Suponha que o estado corrente seja $n = 10$, isto significa que já foram observados $X_0, X_1, ..., X_{10}$. Em geral a previsão do estado do sistema no tempo $n = 11$ dependeria de todos os estados anteriores a ele, porém isto pode ser simplificado se X_{11} depende apenas de X_{10}, de modo que conhecer $X_0, X_1, ..., X_9$ se torna redundante uma vez que X_{10} é conhecido. Se o sistema tem essa propriedade para todo o tempo n, tem propriedades Markovianas.

Seja S o conjunto de valores que Xn pode tomar para qualquer n. Então S é chamado o espaço de estados do processo estocástico $\{Xn, n \geq 0\}$.

Cadeias de Markov de tempo discreto (CMTD) podem ser definidas como processos estocásticos $Xn, n \geq 0$ num espaço de estados S, se para todo i e j em S

$$P(X_{n+1} = j/X_n = i, X_{n-1}, ..., X_0) = P(X_{n+1} = j/X_n = i). \quad (1)$$

Desta fórmula entende-se que o estado futuro X_{n+1} depende somente do estado presente X_n e é independente de seu passado $(X_0, X_1, ..., X_{n-1})$. A probabilidade $P(X_{n+1} = j / X_n = i)$ é chamada de probabilidade de transição de um passo da CMTD no tempo n. Uma CMTD $X_n, n \geq 0$ é dita homogênea, se para todo $n = 0, 1,$

$$P(X_{n+1} = j/X_n = i) = P(X_1 = j/X_0 = i)\,. \quad (2)$$

Em palavras, para todo tempo n a probabilidade de transição de um passo apenas depende de i e j e é sempre igual, portanto, é homogênea no tempo. Neste estudo serão consideradas as CMTD homogêneas e com espaço de estados finitos $S = \{1, 2, ..., N\}$.

Para um passo ($p_{i,j}$) há N^2 probabilidades de transição. Estes valores são organizados em uma matriz *NxN*, esta matriz *P* é chamada de matriz de probabilidade de transição da CMTD. Em que as linhas correspondem ao estado inicial e as colunas ao estado final da transição, assim p_{32}, corresponde a probabilidade de ir do estado *3* para o estado *2* e é armazenado na linha *3* e coluna *2*. Com CMTD pode-se calcular a distribuição de probabilidades dos estados em tempos posteriores, *n* passos:

$$a(n) = a * P^{(n)} \quad (3)$$

sendo $P^{(n)}$ a matriz de probabilidade de transição de n passos, $a = \{a_1, ..., a_N\}$ é a distribuição inicial para $a_i = P(X_0 = i)$, $1 \leq i \leq N$. $a(n) = \{a_1, ..., a_N\}$ é a distribuição prevista para $a_i = P(X_n = i)$, $1 \leq i \leq N$. Considerando $P^{(n)} = P^n$ e P^n sendo a enésima potência da matriz P.

Portanto para fazer a previsão do passo seguinte a equação (3) fica

$$a(1) = a * P \quad (4)$$

3 Estado da Arte

Em [8] é apresentada uma revisão que leva em consideração 150 artigos científicos que usam técnicas convencionais (ARIMA, ARMA, GARCH etc) para realizar previsão de mercado de ações em suas diferentes formas. Os autores afirmam que é muito difícil estabelecer qual é o melhor método, uma das razões é o uso de diferentes medidas de desempenho como erro absoluto médio (MAE), erro previsão absoluto médio (MAPE), desvio absoluto médio (MAD), etc. Afirmam também que somente 63 dos 150 artigos proporcionaram conclusões definitivas sobre qual técnica é a melhor. Além disso, observam que as técnicas tradicionais vão continuar sendo ferramentas promissoras para pesquisa futura.

Em [9] foi publicada uma revisão com 100 artigos científicos que leva em consideração somente os estudos publicados usando redes neurais e redes neurais junto com lógica *fuzzy*. Ilustra diferentes medidas de desempenho adotadas como MAE, MAPE, AIC, etc. Também relata a inclusão de taxa de acerto que mede a porcentagem de predições corretas do modelo. Eles concluem que essas técnicas na maioria dos casos realizam melhor previsão quando comparadas às técnicas convencionais. Mas apontam a dificuldade de definir a estrutura do modelo como número de camadas, além de afirmar que definir a estrutura das camadas representa um procedimento de tentativa e erro que demanda muito tempo.

Em [10] foi realizada uma revisão bibliográfica que considera somente soluções utilizando técnicas de *deep learning*. Segundo os autores, os tipos de redes neurais mais utilizadas são DMLP (Deep Multilayer Perceptron), CNN (*Convolutional Neural Network*) e RNN (*Recurrent Neural Network*) e dentro desta destaca-se a LSTM (*Long-Short Term Memory*). Embora assegurem que soluções utilizando este tipo de técnicas apresentam melhor desempenho, também salientam a existência de soluções híbridas utilizando diferentes técnicas como cadeias de Markov, algoritmos genéticos, cadeias de Markov ocultas, lógica *Fuzzy*. Também apontam que a perspectiva futura é obter soluções que juntem diferentes técnicas.

Em [11] é apresentado uma revisão que leva em consideração técnicas de aprendizado de máquina, modelo de cadeias de Markov ocultos, ARIMA e outras técnicas de *deep learning*. A principal conclusão é que as soluções que consideram uma combinação de métodos diferentes apresentaram melhores resultados.

O componente estocástico do problema de previsão de tendências e preços no mercado financeiro tem motivado o uso de modelos baseados em cadeias de Markov [3,12-15]. Como relatado acima, nenhuma solução atual faz uma previsão exata, portanto soluções usando cadeias de Markov discretas são também um campo a ser explorado [3-5,11]. Além disso, por ser um método probabilístico, não proporciona informação de preço exato, portanto pode ser usado como um método a ser combinado com outras técnicas, para ajudar a tomar decisões [1].

Cadeias de Markov de tempo discreto têm sido amplamente utilizadas para realizar principalmente previsão de tendência de preços de índices e ações [2-6,13,16-18]. Os autores fazem a verificação da previsão que na maioria dos casos foi acertada e recomendam o uso desta técnica. A seguir a descrição de alguns destes artigos.

Em [1] utilizou-se uma modelagem de CMTD considerando 3 estados, sendo estes: a probabilidade de o preço aumentar (preço aumenta), diminuir (preço diminui) ou de permanecer igual. Os autores consideraram 27 dias de variações da ação *China Merchants Banks* de Xangai, levando em consideração o período de 5 de janeiro de 2007 ao dia 12 de fevereiro de 2007. Uma matriz de transição *3x3* foi construída a partir dos preços de fechamento da ação durante

os 27 dias observados. A partir da Eq. (3) foram feitas as previsões para os dias 28, 29 e 30 cujos estados com maior probabilidade de transição foram respectivamente: preço aumenta, preço diminui e preço diminui. Os 3 resultados, foram confrontados com dados reais e confirmaram as previsões.

Semelhante ao que foi feito em [1], em [4] também se implementa um método que considera previsões em 3 estados, porém o método foi aplicado na ação do *State Bank of India (SBI)* da bolsa de valores indiana. Foi considerado um período de movimentos que vai de 21 de março de 2011 até 20 de março de 2015, totalizando 1035 dias utilizados para a construção do conjunto de treinamento. Construiu-se uma matriz de transição *3x3* e com base na Eq. (3) foram apresentadas as previsões para o 1036º dia e para o 1037º dia. O resultado destas previsões também foi de sucesso ao serem comparados aos dados reais.

Em [5], os autores implementam a modelagem de CMTD considerando 3 estados, da mesma maneira como feito em [1]. Neste caso a modelagem foi aplicada na ação *Safaricom Kenya limited* da bolsa de valores da África, foram considerados 784 dias de movimento no período de 1 de abril de 2008 até 30 de abril de 2012. Foi construída a matriz de probabilidade de transição de *3x3* com base nos preços de fechamento da ação no período de 784 dias. Utilizando a Eq. (3) apresenta-se a previsão dos dias 785 e 786, estes resultados são comparados com os dados reais, confirmando a previsão.

Em [3], os autores também trazem uma implementação de CMTD baseada em 3 estados, desta vez aplicado na ação do *Habib Bank Limited (HBL)* da bolsa de valores do Paquistão. Foi construída uma matriz de transição 3x3 considerando os preços de fechamento da ação em 1723 dias de movimento no período de 24 de setembro de 2007 até 20 de fevereiro de 2015. Foram previstos os dias de número 1724 e 1725 utilizando a Eq. (3) e a previsão foi confirmada com base nos dados reais.

Os trabalhos observados utilizaram de 24 a 1723 dias para construção da matriz de probabilidade de transição. Além disso, foram utilizados poucos dias para determinar a exatidão da previsão. Neste trabalho pretende-se realizar experimentos que possam relacionar a exatidão da previsão em função do período selecionado para construir a matriz de transição de probabilidade, além de considerar mais dias para determinar a exatidão da previsão.

4 Materiais e Métodos

A metodologia aplicada para o desenvolvimento deste trabalho começou pelo estudo detalhado do método baseado em CMTD de 3 estados. Seguidamente foi realizada a seleção e compilação dos dados necessários para o desenvolvimento. Os dados foram extraídos do site da B3 (https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/cotacoes-historicas/), que disponibiliza as cotações diárias de todos os papeis negociados na bolsa de valores brasileira em arquivos de texto agrupados por ano. Foram coletados os dados de 2010 até 2021. Para este fim usou-se a biblioteca *libcurl* do PHP que permite a conexão e comunicação com a API da bolsa de valores de São Paulo que fornece os dados através do protocolo HTTP (protocolo de transferência de hipertexto). Os dados históricos obtidos a partir destas consultas foram armazenados no banco de dados MongoDB, que é um banco de dados NoSQL (ou não-relacional) [19]. Depois utilizando um código escrito na linguagem Python com auxílio da biblioteca Pymongo foram identificadas quantas e quais ações tinham registros do ano de 2010 à 2021 e foram identificadas no total 313 ações. Para realizar a análise quantitativa deste trabalho foram consideradas 50 destas ações, o que representa um pouco mais de 15% do número total de ações. Essas ações foram selecionadas a partir de 5 índices setoriais da B3, que agrupam empresas de um mesmo setor. De cada um desses índices selecionou-se 10 ações levando em consideração o maior volume de participação no mercado. As ações foram selecionadas dos índices IEE, INDX, IFNC, IMOB e ICON correspondentes aos setores de energia elétrica, indústria, financeiro, imobiliário e comercio respectivamente.

A seguinte etapa da metodologia foi a implementação de método de três estados, construindo a matriz de transição que será explicada no item desenvolvimento. A implementação foi realizada em *PHP* utilizando o framework *YII* e permite a seleção de qualquer ação, das 131, do período para construir a matriz de transição e realiza a previsão do dia seguinte.

Uma vez que a ferramenta foi implementada e verificado o correto funcionamento da mesma procedeu-se para a análise de como o período considerado na construção da matriz de transição afeta a previsão realizada. Para esse fim foram calculadas as matrizes de transição para cada ação em que os períodos de tempo considerados para a construção dessas matrizes foram de 3, 6 e 9 anos. Portanto, com o objetivo de verificar o desempenho da exatidão do método implementado foi verificado se a tendência prevista ficava conforme com a tendência real. Inicialmente a comparação entre a tendência real e a prevista foi repetida durante todo o ano de 2019, sendo que a cada previsão uma nova matriz de transição foi calculada deixando o período de tempo (construção da matriz) do mesmo tamanho, criando uma espécie de "janela de deslizamento", e o número de sucessos e erros foram acumulados, podendo então calcular a percentagem de sucessos e erros durante o ano de 2019. A mesma metodologia foi aplicada para os anos 2020 e 2021. Foi feita uma comparação dos resultados obtidos nestes anos a fim de observar se porventura haveria alguma mudança no

comportamento dos métodos para cada ano, uma vez que em 2019 a pandemia do COVID-19 não estava tão presente no mundo como em seus anos subsequentes.

A seguinte etapa consistiu em propor estratégias de investimento que utilizam a previsão fornecida pelo método e verificar o efeito nos investimentos.

5 Desenvolvimento

Como o nome já sugere, o método de 3 estados realiza previsões com base em três possíveis estados para os preços de fechamento das ações sendo que o primeiro estado Q_1 contém a contagem dos registros quando os preços subiram, Q_2 a contagem dos registros dos preços que permaneceram iguais (com o mesmo valor) e Q_3 a contagem dos registros dos preços que desceram. Sendo assim, dado um período T, seu primeiro registro é considerado como sendo parte de Q_2 e daí em diante os demais registros serão classificados com base no estado do dia anterior, de modo que se o preço de fechamento do primeiro dia está em Q_2, se no segundo dia o preço de fechamento subir ele será contado em Q_1, do contrário se ele descer será contado em Q_3, porém se o preço do segundo dia for igual ao preço do primeiro dia ele será listado em Q_2.

Após, classificação de cada registro em seu devido estado a matriz de transição de probabilidade *3x3* é montada, o cálculo utilizado para contabilizar as transições, dado que Q_{ij} armazena o número de transições do estado i para o estado j e Q_i armazena o número de registros no estado i, a probabilidade de transição de um elemento sair do estado i e passar para o estado j é calculada como $p_{i,j} = Q_{ij}/Q_i$.

Também para obtenção do vetor de estado inicial do sistema é observado o estado do preço de fechamento do último dia do período, o que retornará um vetor de 3 posições, cuja posição que recebe valor *1* representa o estado em que o dia observado se encontra, as demais posições recebem o valor *0*.

Uma vez tendo a matriz de probabilidade de transição (P) e o vetor de estado inicial (a), pode-se fazer a previsão do dia seguinte utilizando a Eq. (4). A previsão proporciona um vetor com três valores de probabilidade, correspondentes a cada um dos 3 estados. Escolhe-se como resposta o estado que tenha a maior probabilidade. Esta resposta então é comparada com a tendência real, se está acertada é marcada como sucesso do contrário será marcada como erro. Isto será feito durante todo o ano 2019, 2020 e 2021 como descrito na metodologia.

Também foram implementadas três estratégias de compra e venda, com o objetivo de estabelecer uma relação entre a qualidade da previsão da tendência e o lucro obtido pela estratégia. Em todas as estratégias, ficou estabelecido que o investidor começa com um total de R$100 e realiza suas transações de compra e venda de acordo com as regras de cada estratégia. Ao final do período de validação, o saldo em ações é convertido em saldo em dinheiro, multiplicando o valor da ação naquele dia pelo número de ações detidas, para simplificação de cálculos não foram considerados os gastos com a corretagem das operações. Sendo que uma das três estratégias é a estratégia de *buy and hold*, na qual o investidor compra o máximo de ações possível com o valor disponível e vende todas as ações ao final do período, sem realizar mais nenhuma outra movimentação. A Estratégia 1 (*E1*) consiste em: Se o investidor tiver dinheiro disponível e a previsão disser que o preço irá subir ou manter: Ele deve comprar o máximo de ações possíveis. Se o investidor tiver ações em sua carteira e a previsão disser que o preço irá baixar ou manter: Ele deve vender todas suas ações. A Estratégia 2 (*E2*) consiste em: Se o investidor tiver dinheiro disponível e a previsão disser que o preço irá subir ou se manter; Ele deve comprar o máximo de ações possíveis. Se o investidor tiver ações em sua carteira: Ele deve vender todas suas ações, se o preço da ação atual x quantidade de ações for maior que R$ 100,00, independente da tendência da previsão do preço.

6 Resultados

Na Figura 1 mostra-se a tela na qual o usuário faz o pedido para previsão do dia seguinte. Ele define o nome da ação, o período de construção da matriz de transição com data inicial e final.



Figura 1: Tela para fazer previsão do dia seguinte para uma ação específica.

A Figura 2 mostra a resposta da ferramenta para previsão do dia seguinte.

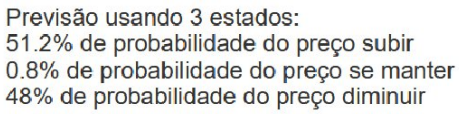


Figura 2: Resposta da previsão do dia seguinte.

A ferramenta também permite realizar uma análise de previsão por um período de tempo determinado. A Figura 3 mostra essa tela.

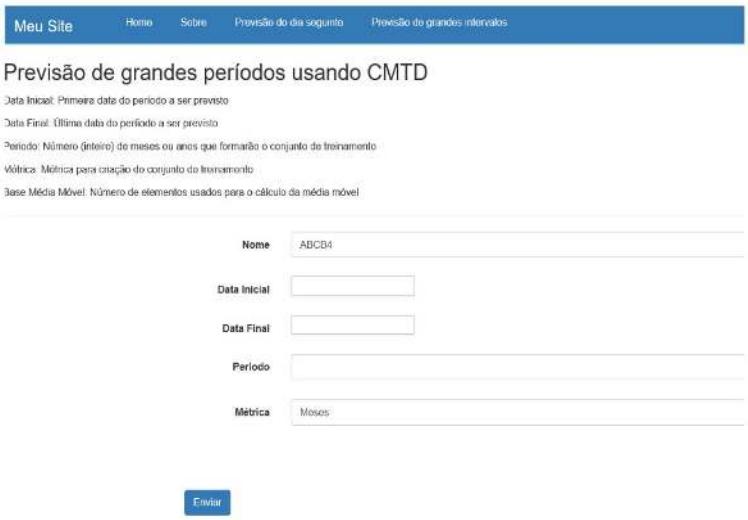


Figura 3: Previsão para um determinado período de tempo, informando também o período da matriz de transição.

A Figura 3 mostra que o usuário escolhe inicialmente uma ação específica, depois entra com data inicial e final do período que quer medir a percentagem de acertos. Em seguida define o número de meses ou anos que a matriz de transição vai ser construída para essa análise e na métrica seleciona se é anos ou meses respectivamente. A Figura 4 mostra o resultado do pedido feito, nesse caso a ferramenta emite um resultado similar ao ilustrado na Fig. 2.
Além disso a ferramenta permite observar o comportamento da ação durante o intervalo analisado, como ilustrado na Fig. 5.



Figura 4: Resposta ao pedido de percentagem de acerto para um período específico.

Além disso, a ferramenta permite observar o comportamento da ação durante o intervalo analisado, como ilustrado na Fig. 5.

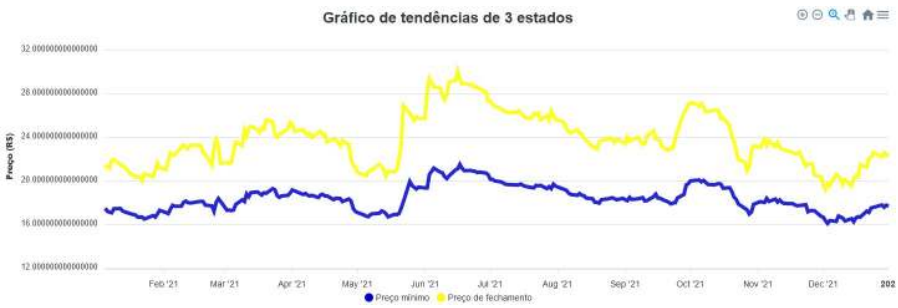


Figura 5: Preço de fechamento e preço mínimo de uma ação selecionada durante um período de tempo determinado.

A Figura 5 ilustra o preço de fechamento em amarelo e o preço mínimo em azul durante o período de verificação da previsão, para uma ação específica.

Com o objetivo de alcançar uma robustez estatística foram levadas em consideração dados financeiros das 50 ações mencionadas anteriormente durante os anos 2010 até 2021. Inicialmente foram construídas matrizes de probabilidade de transição considerando períodos de 3, 6 e 9 anos para cada ação. Levando em consideração estas matrizes, foi realizada a previsão durante o ano 2019, depois durante o ano 2020 e finalmente durante o não 2021. Depois foi calculada a porcentagem de acertos de previsão de cada ação por ano, de 2019 a 2021, utilizando a técnica descrita na metodologia.

Na Tabela 1 mostra-se a média total e seu respectivo desvio padrão das porcentagens de sucesso durante um ano para o conjunto das 50 ações acima mencionadas nos anos 2019, 2020 e 2021, com relação ao período de construção da matriz de transição considerando 3, 6 e 9 anos. Esta Tabela ilustra uma aparente ausência de um padrão consistente nos resultados coletados, uma vez que não há uma relação clara entre o período de construção da matriz de transição e os resultados obtidos. Notou-se que, para o ano de 2019, a maior taxa de sucesso ocorreu com um período de treinamento de 3 anos, seguido por 9 e 6 anos. Já para o ano de 2020, o melhor resultado foi obtido com 6 anos de treinamento, seguido de 3 e 9 anos. Em relação a 2021, o melhor resultado foi alcançado com um período de treinamento de 9 anos, seguido por 3 e 6 anos. Isso sugere que a precisão da previsão não é sensível ao tamanho do período de anos considerado para a construção da matriz de transição. Levando em consideração a média de percentagem de acertos envolvendo os anos 2019 até 2021, observa-se que o período de 3 anos foi o de maior média e menor desvio padrão (49,83%, 0,28).

Portanto, para nossa análise utilizamos o menor período disponível, pois é a melhor relação de custo/benefício, pelo tempo de execução menor com relação à construção da matriz de transição. Dos resultados obtidos nota-se que houve pouca diferença nas porcentagens de previsão quando comparadas entre os anos 2019 até 2021, o que leva a concluir que este método não alterou suas previsões com relação à COVID 19.

Tabela 1: Média total e desvio padrão das porcentagens de acerto por ano e sua relação com o período de construção da matriz de transição.

Ano de validação	Período considerado para construção da matriz de transição em anos					
	3 anos		6 anos		9 anos	
	Média Total	Desvio Padrão	Média Total	Desvio Padrão	Média Total	Desvio Padrão
2019	49,61	3,62	48,47	3,35	49,08	3,50
2020	50,15	4,16	50,46	3,27	49,98	3,38
2021	49,74	4,70	49,13	4,40	50,23	4,64
	49,83	0,28	49,35	1,01	49,76	0,60

A Tabela 2 apresenta uma análise mais detalhada dos resultados obtidos para as 10 ações com os maiores percentuais de sucesso durante os anos de 2019, 2020 e 2021 e um período de 3 anos para construção da matriz de transição. Nessa Tabela observa-se que a média da porcentagem de acertos na previsão para essas 10 ações durante os anos 2019, 2020 e 2021 foi de 52,72% com desvio padrão de 2,29%. Também pode-se observar que a média da porcentagem de acertos no ano de 2020 e 2021 foi de 55,22% e 52,22% respectivamente e no ano de 2019 foi de 50,73%.

Tabela 2**:** Ações com as melhores porcentagens de sucesso com período de construção da matriz de 3 anos

Ação	2019	2020	2021	Média Total	Desvio Padrão
GGBR4	49,60	59,44	45,75	51,59	7,06
WEGE3	51,61	58,63	44,13	51,45	7,25
ODPV3	49,60	51,81	57,89	53,10	4,29
BRFS3	49,60	51,41	57,49	52,83	4,13
SANB11	46,77	53,82	57,09	52,56	5,27
ENGI11	54,03	53,01	56,68	54,37	1,56
RENT3	51,61	56,22	48,99	52,27	3,66
GOAU4	50,81	56,22	47,37	51,46	4,46
MULT3	47,98	55,82	52,63	52,14	3,94
CIEL3	55,65	55,82	54,25	55,24	0,86
Média Total	50,73	55,22	52,22	52,72	2,29

Uma observação é que o comportamento da previsão depende do ano, ou seja, a depender do ano, o método pode ser mais eficiente ou não, fato não relatado em publicações referentes ao tema e que em trabalhos futuros pretende-se explorar para sintonizar o tamanho do período para construir a matriz de transição ao longo dos diferentes anos. Outra observação é que a média das previsões foi maior que 50% para esse caso de uso.

Pelos resultados obtidos, tem-se que considerando as 50 ações durante os anos de 2019 até 2021 a média da percentagem de acertos foi de 49,83%. Ao considerar as 10 ações com melhores resultados a média da percentagem de acertos foi de 52,73%.

A nível nacional tem-se que em [20] apresentaram-se duas soluções baseadas em RNA (redes neurais artificiais), sendo uma das soluções uma RNA pura e outra híbrida, onde obteve-se taxas de acerto de previsão de tendências no índice Ibovespa de 55,04% e 55,74% respectivamente. Foram utilizados algoritmos genéticos e uma base de dados com registros históricos do Ibovespa de 01 de setembro de 2006 até 11 de abril de 2018.

Em [21] foi implementado um método que utiliza diversos algoritmos de aprendizado de máquina (Ensemble) e comparou com os resultados obtidos por algoritmos de redes neurais do tipo Perceptron Multi-Camadas (MLP), Não-linear Auto-regressiva com Entradas Exógenas (NARX) e Memória de Curto e Longo Prazo (LSTM). Os modelos foram treinados com dados de 2010 a 2017. Ao analisar o movimento da previsão para Alta ou Baixa do índice Ibovespa, pôde-se constatar que as redes MLP e NARX, apresentaram 80% de acertos, o modelo Ensemble com 70%, enquanto a LSTM ficou com 60%.

A Ref. [22] traz um estudo de caso sobre os preços da ação da Petrobrás (PETR4) na bolsa de valores nacional. Os autores utilizaram redes MLP para realizar previsões nos preços da ação PETR4. Este trabalho apresenta taxas de acertos acima de 80% em seus experimentos, algumas vezes até mesmo acima dos 90%.

É importante salientar que não se pode fazer uma comparação direta com as diferentes contribuições mencionadas acima pois uns artigos expõem previsão de índices a diferença de um grupo de ações específicas. Além disso, os períodos de treinamento e anos são diferentes. Cabe mencionar também que o trabalho publicado em [22] foi dedicado para uma ação em particular na qual as redes neurais foram especificamente configuradas para esse fim.

Embora a média das porcentagens de acertos para as 50 ações consideradas esteja um pouco abaixo de 50%, observa-se que para ações específicas esta previsão melhorou para perto de 60%. Sabendo que o modelo de cadeias de Markov de tempo discreto pode ser aprimorado como a inclusão de análise de estado estável ou considerando cadeias de maior ordem. Podemos dizer que os resultados obtidos são encorajadores para continuar investindo nesse modelo.

Com o objetivo de analisar como o método pode auxiliar investidores em suas decisões, foram implementadas as três estratégias de investimento explicadas na metodologia. Para este fim, foi selecionado um portfólio composto pelas 10 ações elencadas na Tabela 2. Para comparar os resultados com as estratégias de investimento e estabelecer uma relação entre a qualidade da previsão da tendência e o retorno obtido pela estratégia, os ganhos para o ano de 2019 são apresentados na Tabela 3. Como observado na Tabela 3, no ano de 2019 a estratégia *"buy and hold*" (E3) obteve o melhor resultado, seguido da estratégia 1 (E1) e depois da estratégia 2 (E2). Importante salientar que neste ano a média da porcentagem de acertos usando o método de 3 estados foi a mais baixa de 50,22%, também se observa que mesmo não entregando o melhor resultado, as estratégias E1 e E2 tiveram lucro durante o ano de 2019

A Tabela 4 ilustra o mesmo processo para o ano de 2020. Como observado na Tabela 3, no ano de 2019 a estratégia *"buy and hold*" (E3) obteve o melhor resultado, seguido da estratégia 1 (E1) e depois da estratégia 2 (E2). Importante salientar que neste ano a média da porcentagem de acertos usando o método de 3 estados foi a mais baixa de 50,22%, também se observa que mesmo não entregando o melhor resultado, as estratégias E1 e E2 tiveram lucro durante o ano de 2019. A Tabela 4 mostra que no ano 2020, o maior lucro foi obtido usando E1, além disso mostra que a estratégia *"buy and hold"* não teve lucro durante o ano 2020, ademais podemos ressaltar que nesse ano a média da porcentagem de acertos usando o método de 3 estados foi a mais alta de 55,22%.

Na Tabela 5, mostra-se o resultado obtido aplicando as estratégias de investimento para o ano 2021.

Tabela 3: Resultado em reais das estratégias de investimento aplicadas para 10 ações com melhor desempenho durante 2019.

	Estratégias		
Ação	**E1**	**E2**	**E3**
GGBR4	118,04	109,54	119,28
WEGE3	182,20	131,37	176,70
ODPV3	109,88	115,12	120,40
BRFS3	124,26	129,76	140,96
SANB11	96,84	112,66	97,00
ENGI11	130,50	132,68	107,20
RENT3	130,98	131,38	143,88
GOAU4	123,91	111,44	127,40
MULT3	105,91	118,79	132,16
CIEL3	100,00	100,00	93,94
TOTAL	1.222,52	1.192,74	1.258,92

Tabela 4: Resultado em reais das estratégias de investimento aplicadas para 10 ações com melhor desempenho durante 2020.

	Estratégias		
Ação	**E1**	**E2**	**E3**
GGBR4	196,74	118,36	122,50
WEGE3	230,70	136,57	150,30
ODPV3	86,82	87,55	72,75
BRFS3	53,49	67,54	45,00
SANB11	138,10	93,90	92,42
ENGI11	110,11	113,15	51,37
RENT3	159,72	125,48	134,12
GOAU4	185,65	112,32	113,00
MULT3	100,60	77,92	70,08
CIEL3	100,60	100,00	42,13
TOTAL	1.361,93	1.032,79	893,67

Tabela 5: Resultado em reais das estratégias de investimento aplicadas para 10 ações com melhor desempenho durante 2021.

	Estratégias		
Ação	**E1**	**E2**	**E3**
GGBR4	86,74	104,27	108,52
WEGE3	57,53	57,00	32,74
ODPV3	113,55	83,50	73,74
BRFS3	100,26	100,00	89,44
SANB11	126,82	71,12	60,78
ENGI11	118,19	93,00	43,81
RENT3	92,09	84,32	52,10
GOAU4	89,66	114,54	90,08
MULT3	98,40	91,24	73,64
CIEL3	100,00	100,00	55,75
TOTAL	983,24	898,99	680,60

Na Tabela 5, observa-se que nenhuma das estratégias teve lucro durante o ano de 2022, sendo que a E1 foi a que entregou melhor perda ao investidor. Portanto, observa-se que E1 teve lucro no ano de 2019, a de maior lucro em 2020, além disso em 2021 foi a estratégia que menos perdeu dinheiro. Isto sugere que embora a média da taxa de sucesso das previsões tenha sido de 52,70% ao longo dos 3 anos de avaliação, as previsões podem ajudar o investidor a melhorar suas decisões de investimento.

Importante salientar que pelo algoritmo de investimento das E1 e E2, o investidor não conseguiu encontrar um momento para realizar investimento na ação CIEL3, sendo que em nenhum desses anos foi aplicado o investimento reservado para esta ação. Sabendo que a percentagem de acerto das previsões para CIEL3 nos anos 2019, 2020 e 2021 foram de 55,65%, 55,82% e 54,25% respectivamente, além de verificar que durante esses anos dita ação diminuiu de forma constante, podemos concluir que o conselho de não aplicar dinheiro nesta ação durante esses anos foi uma previsão acertada.

7 Conclusões

Foi desenvolvida uma ferramenta web que permite a previsão de tendência de preço do dia seguinte para qualquer ação da B3. Esta ferramenta permite definir o período de tempo para construir a matriz de probabilidade de transição.

A diferença das publicações relacionadas sobre este tema, nas quais comparam a previsão de acertos com relação a poucos dias, neste trabalho foi comparada com relação a um ano. E a ferramenta permite que o usuário selecione qualquer período que precisar. Também outra diferença é que neste trabalho a análise foi feita para 50 ações e não para poucas ações específicas, como encontra-se na literatura.

Os resultados mostraram que ao considerar 50 ações, a média de previsão de acertos durante o ano de 2019 foi de 49,61%, durante o ano 2020 foi de 50,15% e durante o ano de 2021 foi de 49,74%. Entretanto, também pode-se constatar que para determinadas ações a média de previsão de acertos atingiu até 58,64%.

Os resultados mostrados com relação às estratégias de investimento sugerem que o método pode melhorar os resultados nos investimentos.

Em pesquisa futura espera-se incrementar a percentagem de acertos fazendo análise de estado estável e considerando cadeias de Markov de ordem maior.

Agradecimentos

Agradecimentos para FAPESB e ICB da UESC pelo apoio financeiro em dois projetos de iniciação científica de João Queiroz.

Referências

[1] D. Zhang e X. Zhang, "Study on forecasting the stock market trend based on stochastic analysis method," *International Journal of Business Management*," vol. 4, no. 6, pp. 163–160, 2009. Disponivel em: https://doi.org/10.5539/ijbm.v4n6p163

[2] S. Vasanthi, M. Subha, e S. T. Nambi, "An empirical study on stock index trend prediction using Markov chain analysis," *Journal of Banking Financial Services and Insurance Research*, vol. 1, no. 1, pp. 72–87, 2011. Disponível em: https://www.indianjournals.com/ijor.aspx?target=ijor:jbfsir&volume=1&issue=1&article=005

[3] Q. Sultan, K. Fatima, and J. Ahmed, "Application of Markov Chain to Model and Predict Share Price Movements: A Study of HBL Share Price Movements in Pakistan's Stock Market," *Bi-Annual Research Journal "BALOCHISTAN REVIEW"*, vol. XL, no. 1, 2010, pp. 100–113. Disponível em: https://www.researchgate.net/publication/343650646_Application_of_Markov_Chain_to_Model_and_Predict_Share_Price_Movements_A_Study_of_HBL_Share_Price_Movements_in_Pakistan's_Stock_Market

[4] A. Bairagi e S. Kakaty, "Analysis of stock market price behaviour: a markov chain approach," *International Journal of Recent Scientific Research*, vol. 6, no. 10, pp. 7061–7066, 2015. Disponível em: https://www.recentscientific.com/sites/default/files/3581.pdf

[5] S. Otieno, E. O. Otumba, e R. N. Nyabwanga, "Application of Markov Chain to Model and Forecast Stock Market Trend: a Study of Safaricom Shares in Nairobi Securities Exchange, Kenya," *International Journal of Current Research*, vol. 7, no. 4, pp. 14712–14721, 2015. Disponível em: https://www.journalcra.com/sites/default/files/issue-pdf/8260.pdf

[6] D. N. Choji, S. N. Eduno, e G. T. Kassem, "Markov chain model application on share price movement in stock market," *Computer Engineering and Intelligent Systems*, vol. 4, no. 10, pp. 84–95, 2013. Disponível em: https://iiste.org/Journals/index.php/CEIS/article/view/7972/8139

[7] V. G. Kulkarni, *Modeling and Analysis, Design, and Control of Stochastic Systems*, 1a ed., New York, USA: Springer-Verlag, 1999.

[8] V. Atsalakis, *Computation Optimization in Economics and Finance Research Compendium*, 1a ed., New York, USA: Nova Science Publishers, Inc Editors, 2013.

[9] G. Atsalakis e K. Valavanis, "Survey stock market forecasting techniques – Part II: Soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3(2), pp. 5932–5941, 2009. Disponível em: https://doi.org/10.1016/j.eswa.2008.07.006

[10] O. B. Sezer, M. U. Gudelek, e A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, p. 106181, 2020. Disponível em: https://doi.org/10.1016/j.asoc.2020.106181

[11] R. Sukkati e D. A. Torse, "Stock market forescasting techniques: A survey," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 5, pp. 4842-4844, 2019. Disponível em: https://www.irjet.net/archives/V6/i5/IRJET-V6I5577.pdf

[12] R. Sasikumar e A. S. Abdullah, "Applications of various stochastic models in financial prediction," *International Journal of Scientific and Innovative Mathematical Research (IJSIMR)*, vol. 3, no. 3, pp. 852–857, 2015. Disponível em https://www.researchgate.net/publication/295907256_APPLICATIONS_OF_VARIOUS_STOCHASTIC_MODELS_IN_FINANCIAL_PREDICTION_A_Sheik_Abdullah

[13] N. Redzwan, N. Musa, A. H. A. Latip, Y. A. Latif, e I. N. A. Rahman, "Stock market analysis during election period in Malaysia," *International Journal of Business and Economy,* vol. 1, no. 2, pp. 93-102, 2019. Disponível em: https://myjms.mohe.gov.my/index.php/ijbec/article/view/7828

[14] L. Troiano e P. Kriplani, "Predicting trend in the next-day market by hierarchical hidden Markov model," em *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, Krakow, Poland: IEEE, 2010, pp. 199-204. Disponível em**:** https://doi.org/10.1109/CISIM.2010.5643663

[15] Y. Sun, "Index forecast study based on amended weighted Markov chain in China," *International Journal of Trade, Economics and Finance*, vol. 11, no. 5, pp. 98-103, 2020. Disponível em: https://doi.org/10.18178/ijtef.2020.11.5.674

[16] A. Fritiyanto e T. E. Lestari (2018), "Application of Markov chain to stock trend: A study of PT HM Sampoerna, tbk," em *IOP Conference Series: Materials Science and Engineering, Volume 434, 3rd Annual Applied Science and Engineering Conference (AASEC 2018)*, Bandung, Indonesia: IOP, 2018. Disponível em: https://doi.org/10.1088/1757-899X/434/1/012007

[17] I. Adesokan, "Markov chain asset pricing model for an emerging market", Dissertação de Mestrado, Mestrado em Matemática, Pan African University, Institute for Basic Sciences, Technology and Innovation, Nairobi, Quênia, 2018. Disponível em: http://ir.jkuat.ac.ke/handle/123456789/4694

[18] M. K. Bhusal, "Application of markov chain model in the stock market trend analysis of Nepal," *International Journal of Scientific Engineering Research,* vol. 8, no. 10, pp. 1733-1745, 2017. Disponível em: https://www.ijser.org/researchpaper/Application-of-Markov-Chain-Model-in-the-Stock-Market-Trend-Analysis-of-Nepal.pdf

[19] R. P. Padhy, M. R. Patra, e S. C. Satapathy, "RDBMS to NoSQL: Reviewing some next-generation non-relational databases," *International Journal of Advanced Engineering Science and Technologies*, vol. 11, no. 1, pp. 15–30, 2011. Disponível em: https://www.researchgate.net/profile/Rabi-Padhy/publication/265062016_RDBMS_to_NoSQL_Reviewing_Some_Next-Generation_Non-Relational_Database%27s/links/5476b2930cf29afed61424a6/RDBMS-to-NoSQL-Reviewing-Some-Next-Generation-Non-Relational-Databases.pdf

[20] W. Castello Branco Neto, A. A. Salvi, e W. P. Souza, "Hybrid neural networks applied to Brazilian stock market," *Revista de Informática Teórica e Aplicada*, vol. 7, no. 2, pp. 42–65, 2020. Disponível em: https://doi.org/10.22456/2175-2745.88911

[21] R. C. Dametto, "Estudo da aplicação de redes neurais artificiais para predição de séries temporais financeiras," Dissertação de mestrado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Estadual Paulista, Bauru, Brasil, 2018. Disponível em: http://hdl.handle.net/11449/157058

[22] F. A. Oliveira, C. N. Nobre, L. E. Zárate, "Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil," *Expert Systems with Applications,* vol. 40, no. 8, pp. 7596–7606. Disponível em: https://doi.org/10.1016/j.eswa.2013.06.071