

ARTIGO ORIGINAL

## Previsão de intervalos de preço no mercado de ações brasileiro usando cadeias de Markov de tempo discreto

## Forecasting price intervals in the Brazilian stock market using discrete-time markov chains

Martha Ximena Torres Delgado <sup>1</sup>, João Queiroz <sup>1</sup>, Oniram Átila<sup>1</sup>

<sup>1</sup>Universidade Estadual de Santa Cruz

\*mxtd@uesc.br; johnrick.jh@gmail.com; oniram@gmail.com

Recebido: 11/03/2022. Revisado: 03/11/2022. Aceito: 31/03/2023.

### Resumo

As Cadeias de Markov de tempo discreto tem sido preferencialmente usadas para fazer previsão de tendência de preços de ações e de índices utilizando três estados (preço subir, preço baixar, preço ficar estável) e análise de estado estável. Em quanto que a previsão de intervalos de preço tem sido pouco explorada. Neste trabalho realiza-se a implementação de três maneiras diferentes de construir a matriz de probabilidade de transição para previsão de intervalos de preço, compara-se a previsão com dados reais e mede-se a porcentagem de acertos. Além disso, calcula-se a relação entre a porcentagem de acertos e o período de construção da matriz de transição de probabilidade e também a relação entre a porcentagem de acertos e o número de intervalos de preço ou estados. A análise foi feita utilizando 10 ações aleatórias da bolsa de valores de São Paulo com dados de 2010 até 2019. Um dos métodos avaliados que consistiu em intervalos de tamanho fixo, usando matriz de transição de 12 meses e 5 intervalos, foi o que melhor desempenho apresentou, obtendo média total de porcentagem de acertos acima de 81%. Ademais, quatro estratégias de investimento foram implementadas levando em consideração os resultados deste método de previsão, mostrando que é possível incrementar os investimentos com os resultados do método.

**Palavras-Chave:** Cadeias de Markov de Tempo Discreto; machine learning; previsão do mercado de ações.

### Abstract

Discrete-time Markov Chains have been preferentially used to predict the trend of stock and index prices using three states (price rise, price fall, price stay steady) and steady state analysis. While the prediction of price intervals has been little explored. In this work, we implement three different ways of building the transition probability matrix for forecasting price intervals, this prediction is compared with actual data and the hits percentage is measured. In addition, the relationship between the hits percentage and the period of construction of the probability transition matrix is calculated, as well as the relationship between the percentage of hits and the number of price intervals or states. The analysis was performed using 10 random stocks from São Paulo stock exchange with data from 2010 to 2019. One of the methods evaluated, which consisted of fixed-length intervals, using a transition probability matrix of 12 months and 5 intervals, was the one that presented the best performance, obtaining a total average of hits percentage above 81%. Furthermore, four investment strategies were implemented taking into account the results of this method, showing that it is possible to increase investments with the results of the method.

**Keywords:** Discrete-Time Markov Chains; machine learning; stock market prediction.

## 1 Introdução

Investidores em geral tem mostrado forte interesse em ferramentas que ajudem a prever a tendencia do mercado de ações. Estas ferramentas procuram prever valores de índices e preços de ações com maior precisão possível. Existe grande quantidade de estudos implementando diferentes métodos como análises de regressão, ARIMA (Autorregressive Integrated Moving Average), data mining, redes neurais, cadeias de Markov ocultas e cadeias de Markov de tempo discreto (Atsalakis and Valavanis, 2009; Sukkatti and Torse, 2019). Entretanto, a previsão de índices e preços de ações é uma tarefa difícil dada a volatilidade do mercado que é influenciada por diversos fatores que envolvem condições políticas, sociais, econômicas globais e locais além de muitos outros fatores. Portanto, é consenso geral que o uso de somente um método não é suficiente para prever exatamente mudanças diárias no mercado de ações (Bairagi and Kakaty, 2015; Otieno et al., 2015; Sultan et al., 2019; Sukkatti and Torse, 2019). Este trabalho foca em métodos de previsão baseados em cadeias de Markov de tempo discreto que podem ser usados em combinação com outros métodos.

As cadeias de Markov de tempo discreto têm sido utilizadas principalmente como ferramenta para prever a tendencia do mercado de ações seja através de índices do mercado ou preço de fechamento de ações, utilizando três estados: o preço aumenta, o preço diminui ou permanece igual e calculando o comportamento de estado estável. Os artigos que desenvolvem esta modelagem, seja de índices (Zhang and Zhang, 2009; Vasanthi et al., 2011), ou de preço de fechamento de ações (Sultan et al., 2019; Bairagi and Kakaty, 2015; Otieno et al., 2015; Zhang and Zhang, 2009; Choi et al., 2013; Esbond and Saporu, 2017; Fritiyanto, 2018) afirmam que a informação derivada desta modelagem ajuda aos investidores externos e locais a identificar as tendencias futuras de ações e mercados de ações (índices) servindo como um indicador útil para tomar melhores decisões de investimento. Em outras palavras, proporciona uma melhor visão sobre o comportamento do mercado.

De outro lado, as cadeias de Markov de tempo discreto têm sido pouco exploradas como ferramenta para previsão de valores absolutos de preços de ações. Neste caso, os preços de fechamento das ações são organizados de maior a menor e dividido em intervalos de tamanho fixo (Zhang and Zhang, 2009) ou de tamanho variável através de histograma (Mittra and Riggieri, 2011), onde os estados das cadeias de Markov de tempo discreto seriam esses intervalos conseguindo dar uma previsão sobre o intervalo preço da ação no dia seguinte.

Além disso, é importante salientar que tanto para a modelagem usando três estados (preço aumenta, diminui ou permanece igual) quanto para intervalos de preços de fechamento, a comprovação da exatidão da previsão tem sido feita com máximo três dias futuros.

Portanto, os objetivos deste trabalho são (1) implementar o modelo de cadeias de Markov de tempo discreto usando intervalos de preço de fechamento de ações. Esta modelagem será feita com 10 ações (BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 e VALE3) da bolsa de valores de São Paulo (Bovespa) com dados de 2010 até 2019; (2) avaliar a exatidão da previsão compa-

rando resultados de um ano de movimento financeiro. (3) testar diferentes estratégias de investimento utilizando as informações fornecidas pelo método.

O resto do artigo está organizado como segue, no item 2 será relatado o contexto teórico da pesquisa realizada, 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 apresenta-se as conclusões e trabalhos futuros deste trabalho.

## 2 Contexto Teórico

Em seguida apresentam-se os conceitos fundamentais de cadeias de Markov de tempo de discreto baseado em Kul-karni (1999). Considere um sistema que evolui de maneira aleatória no tempo. Suponha que o sistema é observado em tempos  $n = 0, 1, 2, 3, \dots$ . Seja  $X_n$  o estado do sistema em tempo  $n$ , a sequencia de variáveis aleatórias  $\{X_0, X_1, X_2, \dots\}$  é chamado de um processo estocástico de tempo discreto e é escrito como  $\{X_n, n \geq 0\}$ . Seja  $S$  o conjunto de valores que  $X_n$  pode tomar para qualquer  $n$ . Então  $S$  é chamado o espaço de estados do processo estocástico  $\{X_n, n \geq 0\}$ .

Uma cadeia de Markov de tempo discreto (CMTD) é um processo estocástico  $\{X_n, n \geq 0\}$  num espaço de estados  $S$ , se para todo  $i$  e  $j$  em  $S$ ,

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

Em palavras, dado o estado presente do sistema ( $X_n$ ), o estado futuro da CMTD ( $X_{n+1}$ ) é independente de seu passado ( $X_0, X_1, \dots, X_{n-1}$ ). A quantidade  $P(X_{n+1} = j | X_n = i)$  é chamada de probabilidade de transição de um passo da CMTD no tempo  $n$ . Uma CMTD  $\{X_n, n \geq 0\}$  é de tempo homogêneo se para todo  $n = 0, 1, \dots$

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

Isto significa que a probabilidade de transição de um passo depende de  $i$  e  $j$  e é a mesma para qualquer tempo  $n$ , portanto é homogênea no tempo. Neste estudo serão consideradas as CMTD homogêneas e com espaço de estados finitos  $S = \{1, 2, \dots, N\}$ .

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

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

Onde,  $P^{(n)}$  é matriz de probabilidade de transição de  $n$  passos,  $a$  é a distribuição inicial, sendo  $a = \{a_1, \dots, a_N\}$  onde  $a_i = P(X_0 = i)$ ,  $1 \leq i \leq N$ . Sendo  $P^{(n)} = P^n$  e  $P^n$  é a  $n$ -ésima

potencia da matriz  $P$ .

Além disso, com CMTD pode-se calcular as probabilidades de estado estável utilizando as equações de estado estável:

$$\pi_j = \sum_{i=1}^N \pi_i * p_{ij}, j \in S \quad (4)$$

$$\sum_{i=1}^N \pi_j = 1 \quad (5)$$

Onde  $N$  é o número de estados da Cadeia de Markov.

### 3 Estado da arte

Ao longo do tempo, muitos pesquisadores têm-se dedicado ao desenvolvimento de técnicas para previsão do mercado de ações, tal como é assinalado em [Atsalakis \(2013\)](#). Os autores apresentam uma revisão levando em consideração 150 artigos científicos que usam modelos convencionais para realizar previsão de mercado de ações em suas diferentes formas. As técnicas convencionais consideradas foram o processo autorregressivo e de médias móveis integrado (ARIMA), modelo autorregressivo (AR), processo autorregressivo e de médias móveis (ARMA), modelo autorregressivo condicionalmente heterocedástico/heteroscedástico generalizado (GARCH), passeio aleatório, volatilidade estocástica (SV), entre outros. 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. Outros utilizam estimador padronizado de verossimilhança máxima (SML), teste de qui-quadrado, critério de informação de Akaike (AIC), etc. Afirmam também que somente 63 dos 150 artigos proporcionaram conclusões definitivas sobre qual técnica é a melhor. Também cabe salientar que nas conclusões ele relata que soluções utilizando técnicas GARCH em muitas situações foram melhores que os outros métodos, sem excluir situações onde o oposto também ocorre. Também relata que as técnicas tradicionais continuam a ser ferramentas promissoras para pesquisa futura.

Em [Atsalakis and Valavanis \(2009\)](#) foi publicada uma revisão que leva em consideração somente os estudos publicados usando redes neurais e redes neurais junto com lógica fuzzy, foram levados em consideração 100 artigos científicos. Descreve diferentes medidas de desempenho adotadas como MAE, MAPE, AIC, etc. Também relata a inclusão de taxa de acerto que mede a porcentagem de previsões corretas do modelo. Eles apontam que essas técnicas na maioria dos casos realizam melhor previsão quando comparadas às técnicas convencionais. Mas aponta 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.

Uma revisão bibliográfica relatada em [Sezer et al.](#)

(2020), 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. Além disso, apontam que a perspectiva futura é obter soluções que juntem diferentes técnicas incluindo o processamento de linguagem natural entre outras.

Em [Sukkti and Torse \(2019\)](#) é apresentado uma revisão que leva em consideração técnicas de aprendizado de máquina, modelo de cadeias de Markov ocultas, 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 ([Sultan et al., 2019; Abdullah, 2015; Redzwan et al., 2019; Troiano and Kriplani, 2010; Sun., 2020](#)). A importância de soluções usando cadeias de Markov discretas é justificada dado que nenhum modelo baseado somente numa técnica específica vai fazer uma previsão exata ([Bairagi and Kakaty, 2015; Otieno et al., 2015; Sultan et al., 2019; Sukkti and Torse, 2019](#)). 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 ([Zhang and Zhang, 2009](#)).

Cadeias de Markov de tempo discreto tem sido amplamente utilizadas para realizar principalmente previsão de tendência de preços de índices e ações ([Redzwan et al., 2019; Fritiyanto, 2018; Adesokan, 2018; Zhang and Zhang, 2009; Vasanthi et al., 2011; Mitra and Riggieri, 2011; Sultan et al., 2019; Bairagi and Kakaty, 2015; Otieno et al., 2015; Choji et al., 2013; Bhusal, 2017](#)). Os autores fazem verificação da previsão que na maioria dos casos foi correta e recomendam o uso desta técnica.

A seguir serão descritos os principais resultados de estudos realizados com cadeias de Markov de tempo discreto que realizam previsão do dia seguinte e comprovam sua exatidão.

[Zhang and Zhang \(2009\)](#) implementam a modelagem de CMTD considerando os 3 estados mencionados anteriormente: o preço aumenta, o preço diminui ou permanece igual. Neste caso a modelagem, usando [Eqs. \(3\) e \(5\)](#) vai fornecer informações sobre a probabilidade de a ação diminuir, aumentar ou permanecer no mesmo valor. Foram considerados 27 dias de movimento da ação China Merchants Banks em Shanghai durante 5 de janeiro de 2007 até 12 fevereiro de 2007. Foi construída a matriz de probabilidade de transição de 3x3 com base nos preços de fechamento da ação no período de 27 dias. Utilizando a [Eq. \(3\)](#) apresenta-se a previsão do dia 28, 29 e 30. A interpretação dos resultados do vetor de 3 elementos consiste em selecionar o estado com maior probabilidade. Em todos os três casos a confrontação com os dados reais confirmaram a previsão.



Também neste artigo implementa-se a modelagem de CMTD considerando como estados os intervalos do preço de fechamento de uma ação. A ação selecionada foi a Sinopec de Shenzhen, utilizando 24 dias de movimento, iniciando em 31 de janeiro de 2007 até 12 de março de 2007. Os preços de fechamento foram divididos em 6 intervalos com comprimento de 0,25 unidades (tamanho constante) do preço mais baixo até o mais alto. Depois foi construída a matriz de transição de probabilidade considerando os seis estados e utilizando a Eq. (3) foi feita a previsão para os dias 25avo e 26avo, e foram conferidos com os dados reais acertando a previsão.

Bairagi and Kakaty (2015) implementam a modelagem de CMTD considerando 3 estados da mesma maneira como D. Zhang e X. Zhang (2009). Neste caso a modelagem foi aplicada na ação State Bank of India (SBI) da bolsa de valores da Índia, foram considerados 1035 dias de movimento no período de 21 de março de 2011 até 20 de março de 2015. Foi construída a matriz de probabilidade de transição de  $3 \times 3$  com base nos preços de fechamento da ação no período de 1035 dias. Utilizando a Eq. (3) apresenta-se a previsão dos dias 1036avo e 1037avo, estes resultados são comparados com os dados reais, confirmando a previsão.

Otieno et al. (2015) implementam a modelagem de CMTD considerando 3 estados da mesma maneira como D. Zhang e X. Zhang (2009). 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  $3 \times 3$  com base nos preços de fechamento da ação no período de 784 dias. Utilizando a Eq. (3) apresenta-se a previsão dos dias 785avo e 786avo, estes resultados são comparados com os dados reais, confirmando a previsão.

Sultan et al. (2019) implementam a modelagem de CMTD considerando 3 estados da mesma maneira como D. Zhang e X. Zhang (2009). Neste caso a modelagem foi aplicada na ação Habib bank limited (HBL) da bolsa de valores de Pakistan, foram considerados 1723 dias de movimento no período de 24 de setembro de 2007 até 20 de fevereiro de 2015. Foi construída a matriz de probabilidade de transição de  $3 \times 3$  com base nos preços de fechamento da ação no período de 1723 dias. Utilizando a Eq. (3) apresenta-se a previsão dos dias 1724avo e 1725avo, estes resultados são comparados com os dados reais, confirmando a previsão.

Embora Mitra and Riggieri (2011) não apresentem comparação com dados reais, é importante descrever este trabalho porque ele apresenta a modelagem baseada em CMTD utilizando intervalos de preços de ações de maneira diferente ao descrito em Zhang and Zhang (2009). Neste caso a modelagem foi aplicada na ação do Google no período de 2 de setembro de 2009 até 3 de setembro de 2010 dando um total de 263 registros. Neste trabalho os intervalos são de tamanho variável e dependem da construção de um histograma e também do cálculo da média móvel (moving average) dos preços das ações. Neste artigo utiliza-se o conceito de probabilidade de estado estável. Fazem o cálculo do vetor de estado estável usando as Eqs. (4) e (5) para fazer previsão futura do comportamento da ação.

Nestes trabalhos observa-se que o método mais explorado é aquele baseado em CMTD de três estados, onde a previsão é para saber se um índice ou ação vai subir, dimi-

nuir ou ficar igual. Poucas são as abordagens que utilizam intervalos de preço, onde a previsão é para saber em que intervalo de preço a ação ou índice vai ficar, foco principal de nosso trabalho. Os períodos de tempo que foram considerados para construir a matriz de transição de probabilidade variou de 24 dias até 1073 dias. A questão que surge é saber se existe um período ideal para construir a matriz de transição de probabilidade. Além disso foram utilizados poucos dias para determinar a exatidão da previsão. Neste trabalho pretende-se explorar mais dias de previsão e também pretende-se realizar experimentos que possam relacionar a exatidão da previsão em função do período selecionado para construir a matriz de transição de probabilidade.

## 4 Metodologia

O primeiro passo da metodologia aplicada para desenvolvimento deste trabalho foi estudar em detalhe os dois métodos encontrados na literatura que são baseados em CMTD usando intervalos de preço de fechamento. O primeiro método foi baseado em Zhang and Zhang (2009) utilizando intervalos de tamanho fixo, o segundo método foi baseado em Mitra and Riggieri (2011) utilizando intervalos de tamanho variável, histograma e média móvel.

O segundo passo consistiu na seleção e compilação dos dados necessários para o desenvolvimento. Foram selecionadas 10 ações de forma aleatória da bolsa de valores de São Paulo (Bovespa), sendo elas: BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 E VALE3. Foram coletados os dados de 2010 até 2019, utilizados para construção da matriz de transição e para medir a efetividade da previsão. Para este fim usou-se a biblioteca `libcurl` que permite a conexão e comunicação para a bolsa de valores de São Paulo, através do protocolo http. Depois foram recuperadas as informações importantes e armazenadas no banco de dados MongoDB, que é um banco de dados NoSQL (Padhy et al., 2011) escolhido pelo fato dos dados das ações apresentarem melhor desempenho utilizando este tipo de banco de dados.

O terceiro passo foi a implementação de três métodos diferentes para construir a matriz de transição de probabilidade. Estes três métodos foram baseados nos procedimentos descritos em Zhang and Zhang (2009) e em Mitra and Riggieri (2011) e serão explicados no item desenvolvimento. A implementação foi realizada em Python e permite selecionar o período para construir a matriz de transição e também qualquer número de intervalos. As matrizes de transição foram calculadas para cada ação e os períodos de tempo considerados para a construção dessas matrizes foram de 1 mês, 2 meses, 6 meses e 12 meses.

O quarto passo consistiu em implementar a previsão do dia seguinte utilizando a Eq. (3) e com o objetivo de verificar o desempenho da exatidão dos métodos implementados foi verificado o dado real com o intervalo da previsão. Inicialmente, a comparação entre o dado real e a previsão foi repetida durante todo ano de 2011, sendo que a cada previsão uma nova matriz de transição foi calculada, deixando o período de tempo 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 2011.

Desta maneira podemos obter mais valores de previsão para serem comparados com os valores reais, neste caso um ano de valores. Este procedimento foi repetido no ano 2012, calculando a matriz de transição, calculando a previsão e depois realizando a comparação com o dado real. E nos anos subsequentes até 2019, para cada ação.

O quinto passo consistiu em propor diferentes estratégias de investimento que utilizam a informação fornecida pelo método e verificar o efeito nos investimentos.

## 5 Desenvolvimento

Como mencionado anteriormente foram implementados três métodos diferentes para construir a matriz de transição.

### 5.1 Matriz de transição utilizando intervalos de tamanho fixo

O procedimento utilizado para calcular a matriz de probabilidade de transição em [Zhang and Zhang \(2009\)](#) é como segue: os 24 dias de preço de fechamento são divididos em 6 intervalos de comprimento de 0,25 unidades do preço mais baixo até o mais alto. Uma vez estabelecidos os intervalos, então cada registro diário é marcado no intervalo que pertence e se faz a contagem de quantos registros pertence a cada intervalo ou estado ( $Q_i$ ). Também se realiza a contagem da transição sequencial dos registros, por exemplo, se no dia 3 o registro pertence ao intervalo 2 e no dia 4 o registro pertence ao intervalo 3 então conta como uma transição  $Q_{23}$ , assim em  $Q_{ij}$  armazena-se o número de transições  $i$  e  $j$ . Portanto, cada elemento da matriz  $6 \times 6$  calcula-se como  $p_{i,j} = Q_{ij}/Q_i$ .

Depois de calculada a matriz de probabilidade de transição, eles calculam o vetor de estado inicial conferindo o preço de fechamento do 24avo dia e avaliando a que intervalo pertence, por exemplo, se pertencer ao terceiro intervalo então o vetor de estado inicial será  $a = \{0, 0, 1, 0, 0, 0\}$ .

Utilizando a equação [Eq. \(3\)](#) é calculada a previsão que proporciona um vetor de probabilidades, que significa qual é probabilidade do preço do dia seguinte cair em cada um dos intervalos ou estados, o intervalo que proporciona a maior probabilidade é escolhido como a resposta da previsão. Para conferir a consistência da resposta, o intervalo selecionado é comparado com o valor real desse dia, se esse valor se encontra no intervalo selecionado então a previsão é considerada correta.

Na nossa implementação, a matriz de transição é calculada como descrito no parágrafo anterior, além disso permite definir o período utilizado para construir a matriz de transição e o número de estados ou intervalos. Inicialmente é definido o período utilizado para construir a matriz de transição, depois disso então a diferença entre o valor mínimo e o valor máximo do preço de fechamento dividido pelo número de estados proporciona o tamanho fixo do intervalo. A quantidade de intervalos é igual ao número de estados, o intervalo inicial começa com o valor mínimo e vai até  $valormnimo + tamanhofixo$  do intervalo e assim por diante, o preço diário será classificado no intervalo correspondente.

O vetor de estado inicial é calculado pegando o valor de

fechamento da data final do período de construção de matriz de transição e localizando o intervalo a que pertence. Portanto, o vetor de estado inicial, de tamanho número de estados, terá valor 1 na posição correspondente ao intervalo acima mencionado e os demais serão 0.

Depois calculá-se a previsão e compara-se como o dado real como descrito no parágrafo anterior. Esta comparação é feita durante um ano, para os anos 2011 até 2019 como descrito na metodologia.

### 5.2 Matriz de transição de probabilidade utilizando intervalos variáveis, histograma e média móvel

Esta implementação foi baseada no procedimento descrito em [Mittra and Riggieri \(2011\)](#), no qual calcula-se a média móvel (moving average) e a diferença entre o preço da ação e a média móvel é usada para definir os intervalos.  $Mdiamvel - n$  está definida como o cálculo da média dos  $n$  preços em sequência, ou seja, os  $n - 1$  preços anteriores e o presente. No artigo foram levados em consideração 3 valores diferentes de  $n$  (3, 5, 10) e também foram considerados tanto o preço de fechamento quanto de abertura para calcular a média móvel. Foram definidos 4 intervalos ou 4 estados cujo comprimento foi determinado pela distribuição equitativa do resultado da montagem de um histograma cujo numero de barras é calculado como raiz quadrada do numero de observações; onde os intervalos dos extremos ficam com os valores que tenham as diferenças menores e as mais altas respectivamente, os dois intervalos do meio foram proporcionalmente repartidos. Obtendo 4 intervalos de tamanhos diferentes. Uma vez que os intervalos foram definidos, marca-se cada diferença de preço com o intervalo a que pertence depois se faz a contagem do número de transições de cada intervalo ao seguinte intervalo, por exemplo, se a diferença de preço do dia 115 pertence ao segundo intervalo e a diferença de preço do dia 116 pertence ao terceiro intervalo então é contada uma transição  $Q_{23}$ . Também se faz a contagem do número de elementos que pertencem a cada intervalo ( $Q_i$ ). Assim então se calcula a matriz de transição  $4 \times 4$  onde cada elemento será  $Q_{ij}/Q_i$ .

Também neste artigo mostra-se que o uso de média móvel de 3 resulta em tamanho de intervalos menores dando resultados com informação mais útil. Além disso, é importante salientar que apesar de usar tanto tempo de fechamento como abertura, recomenda o uso do preço de fechamento para as previsões pois o preço de fechamento é o preço que será utilizado no dia seguinte no início das operações.

Na nossa implementação a matriz de transição é calculada como descrito no paragrafo anterior, com algumas particularidades. No nosso caso, calcula-se a média móvel de tamanho 3 dos preços de fechamento, onde realiza-se o valor médio de 3 medidas de preço de fechamento, duas anteriores e a atual. Também é calculada a diferença da média móvel que consiste na subtração do preço atual com respeito à média móvel. Isto é feito pelo período definido para construir a matriz de transição. Depois disso, é construído um histograma, cujo numero de barras é calculado como raiz quadrada do numero de observações,

com base nos valores de diferença da média móvel, depois criam-se intervalos de acordo ao número de estados com quantidade de eventos igualmente distribuídos, sempre que seja possível. Depois de criados os intervalos, a diferença da média móvel diária será classificada no intervalo correspondente. A matriz de transição tem dimensões  $n_{\text{merodeestados}} \times n_{\text{merodeestados}}$ .

O vetor de estado inicial é calculado pegando a diferença da média móvel da data final do período de construção de matriz de transição e localizando o intervalo a que pertence. Portanto, o vetor de estado inicial, de tamanho número de estados, terá valor 1 na posição correspondente ao intervalo acima mencionado e os demais serão 0.

A previsão segundo a Eq. (3) proporciona um vetor de probabilidades, que significa qual é probabilidade de a diferença da média móvel do preço do dia seguinte cair em cada um dos intervalos selecionados. Para saber se este método fez uma previsão correta, escolhe-se o intervalo que deu a maior probabilidade e com base nos limites desse intervalo serão calculados os limites do intervalo do preço de fechamento previsto. Depois se verifica se o preço de fechamento real está no intervalo previsto, se está, consideramos que a previsão teve sucesso, senão consideramos que a previsão errou. Esta comparação é feita durante um ano, para os anos 2011 até 2019 como descrito na metodologia. Neste método o tamanho dos intervalos é variável e depende da média móvel e do histograma.

### 5.3 Matriz de transição de probabilidade utilizando intervalos variáveis e histograma

O terceiro método pretende aproveitar as características dos dois métodos anteriores, neste caso é dispensado o cálculo de média móvel e utiliza-se o histograma baseado no preço de fechamento para definir o tamanho dos intervalos. Primeiro é construído o histograma depois os intervalos são criados de acordo ao número de estados com quantidade de eventos igualmente distribuídos, sempre que seja possível. Depois de criados os intervalos, ao igual que os métodos anteriores, o preço diário será agrupado no intervalo correspondente. O resto dos cálculos são similares aos descritos no método de intervalos fixos. Neste método o tamanho dos intervalos é variável e depende somente do histograma

## 6 Resultados

Com o objetivo de saber como o período que se considera para construir a matriz de transição afeta a qualidade da previsão então inicialmente foram executados os 3 métodos acima mencionados com 5 estados e utilizou-se um mês, dois, seis e doze meses para construir a matriz de transição. Foi calculada a porcentagem de acertos de previsão de cada ação por ano, desde 2011 até 2019, utilizando a técnica descrita na metodologia. Com o objetivo de ilustrar este comportamento e ao mesmo tempo ocupar menos espaço no artigo então foi calculada a média total dessas porcentagens levando em consideração o resultado das 10 ações e todos os anos de 2011 até 2019. Na Tabela 1 mostra-se a média total e seu respectivo desvio padrão das porcentagens de sucesso durante um ano para o conjunto

das 10 ações acima mencionadas nos anos 2011 até 2019, para todos os métodos implementados.

**Tabela 1:** Média total e desvio padrão das porcentagens de acerto por ano, correspondente às 10 ações (BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 E VALE3) durante os anos 2011 até 2019 e sua relação com o período de construção da matriz de transição para os 3 métodos implementados

Intervalos Fixos		
Período Construção da Matriz em meses	Média Total	Desvio Padrão
1	54,23	6,29
2	66,08	7,6
6	78,6	6,18
12	82,34	5,3
Intervalos Variáveis e Histograma		
Período Construção da Matriz em meses	Média Total	Desvio Padrão
1	51,86	5,91
2	65,76	6,05
6	78,6	5,95
12	82,1	6,28
Intervalos Variáveis, histograma e média móvel		
Período Construção da Matriz em meses	Média Total	Desvio Padrão
1	23,06	3,02
2	34,24	3,77
6	42,29	6,24
12	44,41	8,75

No método de intervalos fixos, quanto maior o período utilizado para construir a matriz de transição maior foi a porcentagem de sucesso. Portanto, as porcentagens de sucesso utilizando 12 meses foram maiores que utilizando outros tamanhos para construção da matriz de transição e o desvio padrão foi menor.

No método de intervalos variáveis e histograma, em todos os casos a porcentagem de sucesso aumentou com o aumento do período para construção da matriz de transição, sendo obtidas as porcentagens de sucesso mais altas quando se usou 12 meses, mas o desvio padrão aumentou nesse período.

No método de intervalos variáveis, histograma e média móvel, da mesma maneira que os métodos anteriores obteve melhor porcentagem de sucesso utilizando um período de 12 meses e neste caso o desvio padrão foi aumentando com respeito ao tamanho do período. Este método apresentou uma porcentagem de sucesso baixa comparada com os outros métodos, isto pelo fato de o tamanho dos intervalos serem menores dificultando o acerto na previsão.

Em conclusão, a matriz de transição de 12 meses, foi a que maior porcentagem de sucesso apresentou em todos os métodos. Portanto, adotamos para os seguintes passos de análises a matriz de transição de 12 meses.

Depois de fixar o período para construir a matriz de transição, o passo seguinte foi determinar como o número de estados (intervalos) afeta a exatidão da previsão. Para esse fim, todos os experimentos foram rodados usando 5, 8 e 15 estados (intervalos). Com o objetivo de ilustrar



este comportamento e ao mesmo tempo ocupar menos espaço no artigo então foi calculada a média total dessas porcentagens levando em consideração o resultado das 10 ações e todos os anos de 2011 até 2019. Na **Tabela 2** mostra-se a média total e seu respectivo desvio padrão das porcentagens de sucesso durante um ano para o conjunto das 10 ações acima mencionadas nos anos 2011 até 2019, para todos os métodos implementados.

**Tabela 2:** Média total e desvio padrão das porcentagens de acerto por ano, correspondente às 10 ações (BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTTEX3, ITUB4, PETR4 E VALE3) durante os anos 2011 até 2019 e sua relação com o número de estados selecionados para os 3 métodos implementados

Intervalos Fixos		
Estados	Média Total	Desvio Padrão
5	82,34	5,3
8	74,64	6,7
15	53,63	9,88
Intervalos Variáveis e Histograma		
Estados	Média Total	Desvio Padrão
5	82,1	6,28
8	74,18	6,95
15	48,51	7,44
Intervalos Variáveis, histograma e média móvel		
Estados	Média Total	Desvio Padrão
5	44,41	8,75
8	31,57	8,19
15	24,04	8,9

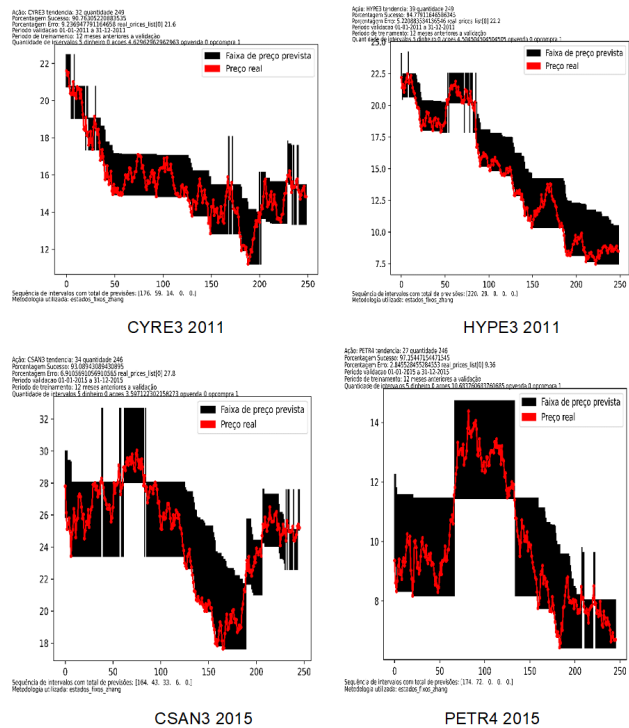
Os resultados apontaram que as melhores previsões foram aquelas que usaram 5 estados em todos os casos. Importante salientar que em todos os métodos ao aumentar o número de estados diminui a porcentagem de sucesso, lembrando que a previsão das cadeias de Markov de tempo discreto proporciona uma probabilidade de um evento ocorrer e não uma certeza absoluta, além disso, a medida que se aumenta o número de intervalos o tamanho dos mesmos vai ficando menor então nesse caso as cadeias de Markov ficam limitadas para determinar qual seria o evento mais provável, levando a maiores erros de previsão.

Além disso, esses resultados apontam que tanto o método de intervalos fixos quanto o método de intervalos variáveis e histograma apresentaram uma média similar sendo que o método de intervalos fixos apresentou desvio padrão menor, considerando 5 estados.

A **Tabela 3** mostra de maneira detalhada os resultados para as 10 ações e todos os anos de 2011 até 2019 para o método de intervalos fixos.

A **Fig. 1** mostra em detalhe a previsão feita para os quatro melhores resultados (CYRE3 2011 (90,76%), HYPE3 2011 (94,79%), CSAN3 2015 (93,09%) e PETR4 2015 (97,15%)) sendo que nesta figura, a curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão.

Como se observa na **Fig. 1**, a ação CYRE3, em 2011, apresenta um comportamento de descida leve e oscilações de preço. Também na **Fig. 1**, a ação HYPE3 em 2011 apresenta



**Figura 1:** Comportamento da previsão das 4 ações com melhores porcentagens de exatidão de previsão, CYRE3 2011 (90,76%), HYPE3 2011 (94,79%), CSAN3 2015 (93,09%) e PETR4 2015 (97,15%). A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão.

um comportamento de descida com pequenas oscilações de preço. Na mesma **Fig. 1**, a ação CSAN3 em 2015 apresenta um comportamento inicial de subida leve com oscilações depois apresenta comportamento de descida e no final oscilações com comportamento de subida. A ação PETR4 no ano 2015, na mesma **Fig. 1**, apresenta inicialmente oscilações depois apresenta um comportamento de subida depois oscilações e depois um comportamento de descida com oscilações. Cabe salientar que na maioria das vezes em que os preços oscilaram, sem nenhuma tendencia, nas 4 ações o método não conseguiu mostrar a tendencia do valor, ele basicamente previu a mesmo intervalo.

Com o objetivo de ilustrar a previsão da tendencia deste método, foi calculado o valor médio de cada intervalo previsto e comparado com o valor real da ação. A **Fig. 2** mostra estes dois valores graficamente para os 4 melhores resultados. Onde a linha azul é o valor médio dos intervalos e a linha vermelha o valor real. Na **Fig. 2**, correspondente a CYRE 2011, algumas oscilações, no início, foram detectadas, além da tendencia de descida. No meio, parte das oscilações não foram detectadas embora a tendencia de descida foi bem ilustrada. Também mostra que no final tenta acompanhar as oscilações. A **Fig. 2**, referente a HYPE 2011, mostra que o método tenta, no início, acompanhar as oscilações depois marca bem a tendencia de descida, mas não consegue acompanhar as oscilações seguintes,

**Tabela 3:** Porcentagem de sucesso da previsão para BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 E VALE3 DTEX3 e ENBR3 usando o método de intervalos fixos (período de construção da matriz de 12 meses e 5 estados) desde 2011 até 2019

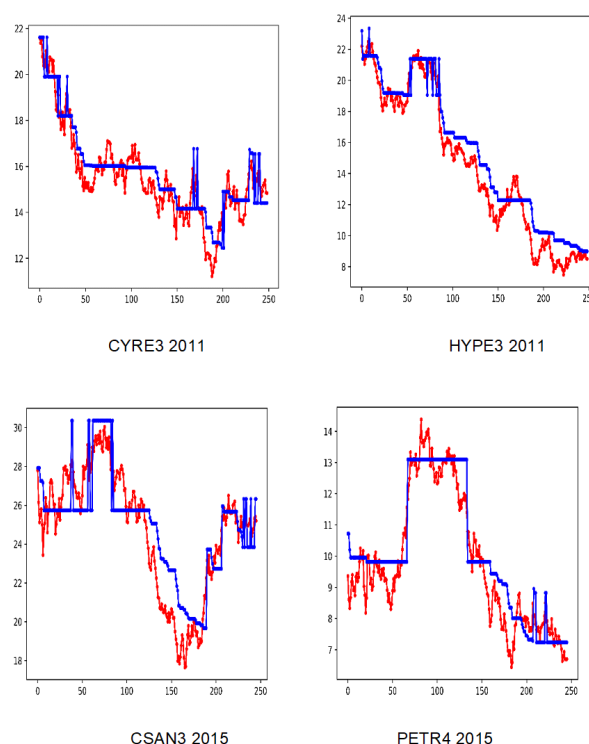
Ação:	2011	2012	2013	2014	2015	2016	2017	2018	2019	Média	Desvio Padrão
BBAS3	84,74	84,96	83,47	79,43	82,11	81,12	80,49	80,41	79,84	81,84	2,09
CSAN3	78,71	77,64	82,26	88,31	93,09	85,94	78,86	83,26	72,98	82,34	6,12
CYRE3	90,76	77,24	77,42	84,68	89,84	83,94	75,61	88,16	77,02	82,74	6,03
ENBR3	69,48	87,8	76,61	79,44	73,58	79,52	74,8	79,18	80,24	77,85	5,1
HYPE3	94,79	82,93	85,89	80,64	80,89	79,92	82,11	82,45	78,22	78,34	5,2
PSSA3	89,56	90,24	80,24	77,42	76,02	89,96	78,86	78,37	81,05	83,09	4,88
DTEX3	81,93	84,15	85,08	89,11	84,15	81,93	84,55	85,71	82,26	84,32	2,3
ITUB4	82,73	82,93	82,26	72,98	80,08	82,73	80,89	86,53	84,27	86,23	5,95
PETR4	89,16	81,3	81,05	81,45	97,15	81,93	87,4	77,96	77,82	81,71	3,76
VALE3	88,75	87,81	82,66	89,52	87,8	83,94	79,27	79,18	84,27	83,91	6,25
Média Total = 82,34 Desvio Padrão = 5,3											

depois acompanha a subida seguinte e tenta acompanhar algumas oscilações para depois acompanhar a tendência de descida, mas sem acompanhar as oscilações de preço. Na Fig. 2, correspondente a CSAN 2015, observa-se que o método não consegue acompanhar muito bem as oscilações de início, depois detecta muito bem as tendências de subida e descida e no final tenta acompanhar as oscilações de preço. No caso da Fig. 2, PETR4 2015, em geral não consegue acompanhar as oscilações de preço mas consegue estimar com certa exatidão a tendência do preço.

A seguir, a Fig. 3 mostra em detalhe a previsão feita para os quatro piores resultados (ENBR3 2011 (69,48%), ITUB4 2014 (72,98%), ENBR3 2015 (73,58%) e CSAN3 2019 (72,98%)), sendo que nesta figura, a curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão.

Como se observa na Fig. 3, a ação ENBR3 em 2011 apresenta muitas oscilações de preço sem uma tendência clara de descida ou de subida. Na Fig. 3 a ação ITUB4 em 2014 também apresenta um comportamento de muitas oscilações de preço, mas com tendências mais definidas de subida e descida de preço. Na Fig. 3 a ação ENBR3 em 2015 apresenta também um comportamento de muitas oscilações, mas com tendências de subida e descida mais claras que a ação anterior. Na Fig. 3 a ação CSAN3 no ano 2019 apresenta muitas oscilações de preço com uma tendência clara de aumento de preço ao longo do ano. Importante observar que nestes casos as oscilações de preço são pequenas e frequentes, em alguns períodos de tempo essas oscilações não foram detectadas enquanto que em outros as oscilações foram detectadas.

Com o objetivo de ilustrar a previsão da tendência deste método, foi calculado o valor médio de cada intervalo previsto e comparado com o valor real da ação. A Fig. 4 mostra estes dois valores graficamente para os 4 piores resultados. Onde a linha azul é o valor médio dos intervalos e a linha vermelha o valor real. Como mostra a Fig. 4, ENBR3 2011, muitas oscilações foram detectadas ao igual que a tendência de subida e descida. Em alguns períodos de tempo as oscilações não conseguiram ser detectadas. A Fig. 4, ITUB4 2014, mostra que o método acompanha algumas oscilações, principalmente na parte final do período, na parte inicial algumas oscilações não conseguiram ser detectadas, mas acompanha as tendências de descida e subida

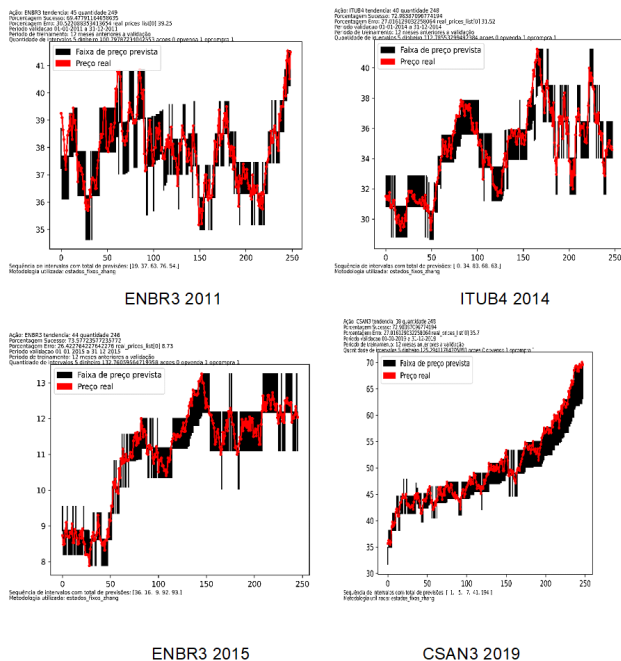


**Figura 2:** Tendência da previsão das 4 ações com melhores percentagens de exatidão de previsão, CYRE3 2011 (90,76%), HYPE3 2011 (94,79%), CSAN3 2015 (93,09%) e PETR4 2015 (97,15%). Linha azul é a média dos intervalos previstos e linha vermelha o preço da ação real

de preço da ação. Na Fig. 4, ENBR3 2015, observa-se que o método tenta acompanhar as oscilações sendo que em alguns trechos não consegue fazê-lo corretamente e também acompanha bem a tendência de descida e subida da ação. No caso da Fig. 4, CSAN3 2019, o método acompanha em alguns trechos as oscilações apresentadas, mas sempre acompanhou a tendência da ação de aumentar seu preço.

Da análise anterior, pode-se concluir que quando a ação





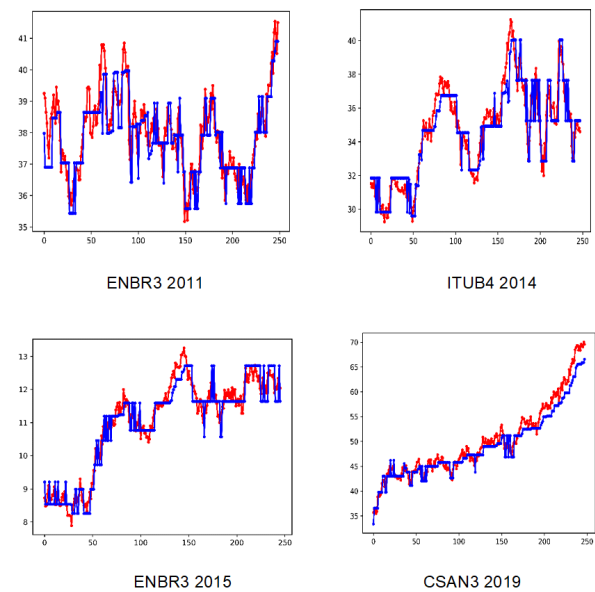
**Figura 3:** Comportamento da previsão das 4 ações com piores percentagens de exatidão de previsão ENBR3 2011 (69,48%), ITUB4 2014 (72,98%), ENBR3 2015 (73,58%) e CSAN3 2019 (72,98%). A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão

apresenta flutuações maiores sem tendencia definida, o método de previsão apresenta % de acertos menores, como mostrado nas Fig. 3 e Fig. 4. O método fica mais assertivo quando as ações apresentam pequenas oscilações e/ou com tendencias definidas, neste caso o método tende a prever o mesmo intervalo anterior, aumentando suas percentagens de acerto, como mostrado nas Fig. 1 e Fig. 2. Mostra-se que a exatidão na previsão depende do comportamento de cada ação. Mesmo com uma porcentagem de previsão menor, 69%, mostra-se que o método consegue prever a tendencia de maneira geral. Na seção seguinte ilustra-se como utilizar os resultados obtidos.

### 6.1 Comparação com outras técnicas

O que este artigo traz é a análise da técnica de cadeias de Markov discreta para previsão de intervalos de preço e não de tendencia, também traz um respaldo estatístico que falta nas demais publicações. Como ilustrado na Tabela 3 a porcentagem de acertos desta solução para as ações consideradas foi de 82,34%.

Comparando este resultado com outros resultados a nível nacional tem-se que em Neto et al. (2020) apresentaram-se duas soluções baseadas em RNA, sendo uma das soluções uma RNA pura e outra híbrida, o autor obteve taxas de acerto de previsão de tendências no índice Ibovespa semelhantes sendo 55,04% e 55,74% respectivamente. Foram utilizados algoritmos genéticos e uma base de dados com registros históricos do Ibovespa de 01



**Figura 4:** Tendência da previsão das 4 ações com melhores percentagens de exatidão de previsão ENBR3 2011 (69,48%), ITUB4 2014 (72,98%), ENBR3 2015 (73,58%) e CSAN3 2019 (72,98%). Linha azul é a média dos intervalos previstos e linha vermelha o preço da ação real

de setembro de 2006 a 11 de abril de 2018.

Dametto (2018) implementou um método que utiliza de 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% e enquanto a LSTM ficou com 60%.

de Oliveira et al. (2013) traz em seu trabalho um estudo de caso sobre os preços da ação da Petrobrás (PETR4) na bolsa de valores nacional. O autor utilizou 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%. Lembrando também que nas redes neurais é necessário configurar a estrutura em cada intervalo de tempo considerado e para cada tipo de ação, não é o caso das cadeias de Markov de estado discreto.

Embora não se possa fazer uma comparação direta com outras técnicas pode-se observar que a presente abordagem pode contribuir para melhorar a taxa de acertos da previsão, em pesquisas futuras espera-se explorar maneiras de juntar diferentes técnicas com o intuito de melhorar a taxa de acertos e facilitar a construção dos modelos.

## 6.2 Estratégias de Investimento

Com o propósito de explorar como o método pode ser usado para melhorar os investimentos serão consideradas 5 estratégias de investimento diferentes, uma delas é uma estratégia padrão que serve como parâmetro de comparação e as outras 4 vão utilizar as previsões obtidas. As condições do investimento serão as seguintes: utilizar um portfólio das 10 ações estudadas, utilizar R\$1.000,00 para comprar ações no início de ano, esse dinheiro será repartido equitativamente para cada ação, então iremos aplicar R\$100,00 para cada ação no início de ano, considerar as operações de compra e venda livres de taxa. Utilizando as diferentes estratégias será analisado quanto seria o dinheiro que teríamos no final de ano. A primeira estratégia, E1, é a estratégia *buyandholding* que é simplesmente comprar o número de ações que o dinheiro disponível permitir, no início de ano e manter as ações até o último dia de ano onde é realizada a venda. Estratégia usada sem ter uma ferramenta de previsão.

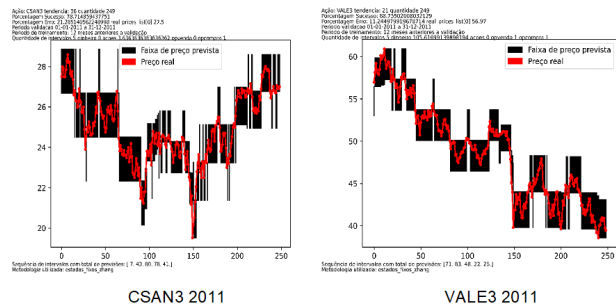
A segunda estratégia, E2, consiste em utilizar o resultado de previsão. Como descrito anteriormente temos um intervalo de valores nos quais o preço da ação pode estar amanhã. Se o preço da ação hoje for menor ou igual ao valor do limite inferior do intervalo previsto então a recomendação é comprar hoje. Se o preço da ação hoje for maior ou igual ao valor do limite superior do intervalo previsto então a recomendação é vender hoje. Se o preço da ação hoje estiver no intervalo previsto não faz nada.

A terceira estratégia, E3, inicialmente funciona da mesma maneira que a anterior, se o preço da ação hoje for menor ou igual ao valor do limite inferior do intervalo previsto então a recomendação é comprar hoje. Se o preço da ação hoje for maior ou igual ao valor do limite superior do intervalo previsto então a recomendação é vender hoje. Neste caso, se o preço da ação hoje estiver no intervalo previsto, então se tenho ações e as comprei por um valor mais baixo que o atual então recomenda-se vender as ações, do contrário, se tenho dinheiro e se o preço da ação hoje estiver na terceira parte inferior do intervalo previsto então recomenda-se comprar ações.

A quarta estratégia, E4, similar à E2 e E3 e no caso de que o preço da ação hoje estiver no intervalo previsto, então se o intervalo anterior for igual ao previsto não vai fazer nada. Se o intervalo anterior estiver numa posição superior ao intervalo previsto então indicativo de baixar o preço portanto vender ações se tiver. Se o intervalo anterior estiver numa posição inferior ao intervalo previsto então a ação vai subir, portanto é hora de comprar.

A quinta estratégia, E5, similar a E4 e no caso de que o preço da ação hoje estiver no intervalo previsto e se o intervalo anterior for igual ao previsto então se tenho ações e as comprei por um valor mais baixo que o atual então recomenda-se vender as ações, do contrário, se tenho dinheiro e se o preço da ação hoje estiver na terceira parte inferior do intervalo previsto então recomenda-se comprar ações.

Importante salientar que E2 e E5 utilizam uma abordagem especulativa, onde se o preço atual estiver na parte inferior do intervalo a probabilidade é que ação pode subir depois. As E4 e E5 tiram proveito do comportamento em “escadinha” das previsões proporcionadas para aju-



**Figura 5:** Comportamento da previsão das 2 ações cujo melhor resultado foi usando E2 no ano 2011: CSAN3 VALE3. A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão

dar na decisão da compra ou venda. A [Tabela 4](#) mostra os resultados de investimento do ano 2011, utilizando as 5 estratégias.

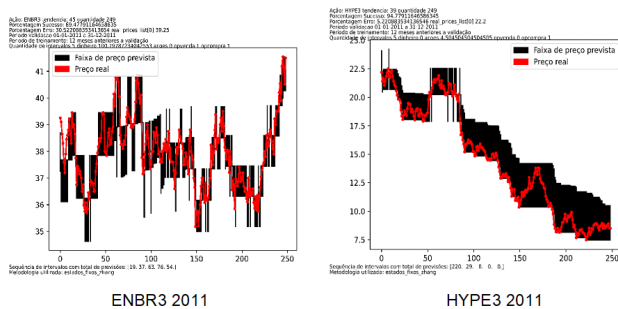
Como foi mencionado anteriormente a E1 corresponde à estratégia buy and holding e serve como parâmetro de comparação com respeito às estratégias que utilizam a informação fornecida pela ferramenta desenvolvida. O ano de 2011 foi um ano particularmente ruim para estas ações. Neste caso, as diferentes estratégias tiveram melhor comportamento que a E1. Mas, isto não pode ser generalizado para os demais anos. Cada ação, em cada ano, apresenta comportamentos diferentes. A [Tabela 4](#) mostra que cada ação tem uma estratégia que tira melhor proveito da informação fornecida pelo método de previsão. Se aplicássemos a estratégia apropriada para cada ação então no final de 2011 teríamos como resultado o valor de R\$ 936,42, um desempenho maior que os obtidos utilizando uma estratégia particular.

Sendo assim, a seguir apresenta-se uma análise para tentar identificar qual seria a melhor estratégia para cada ação. Da [Tabela 4](#) observa-se que CSAN3 e VALE3 tiveram melhor comportamento usando a E2. Como foi descrito anteriormente, E2 somente leva em consideração a previsão quando o preço atual não pertence ao intervalo previsto ou encontra-se exatamente no limite inferior ou superior desse intervalo. A [Fig. 5](#) mostra a previsão para o ano 2011 de CSAN3 e VALE3. Nestes casos o método de previsão de maneira geral não consegue detectar as mudanças diárias de preço, detecta principalmente alguns picos de preço. As oscilações menores de preço não são detectadas. O método de previsão apresenta dificuldade em representar movimentos de descida ou subida através dos intervalos previstos. A solução de maneira geral apresenta intervalos iguais e não intervalos em “escadinha”. CSAN3 da [Fig. 5](#) percebe algum comportamento em “escadinha” mas que não define nenhuma tendência.

Da [Tabela 4](#) observa-se que ENBR3 e HYPE3 tiveram melhor comportamento usando a E3. Como foi descrito anteriormente E3 diferencia-se de E2 porque quando o preço atual pertence ao intervalo previsto, então se tenho ações e as comprei por um valor mais baixo que o atual então recomenda-se vender as ações, do contrário, se tenho dinheiro e se o preço da ação hoje estiver na terceira

**Tabela 4:** Resultado de investimento de 5 estratégias diferentes, utilizando portfólio das ações BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 E VALE3 DTEX3 e ENBR3 no decorrer do ano 2011

Ação:	BBAS3	CSAN3	CYRE3	ENBR34	HYPE3	PSSA3	DTEX3	ITUB4	PETR4	VALE3	TOTAL
E1	76,9	98,5	72,86	104,5	45,2	78,4	57,3	88,62	83,47	82,48	788,33
E2	76,9	98,5	74,38	100,6	42,75	86,93	62,79	88,62	86,31	91,9	809,68
E3	67,52	91,15	91,68	126,77	70,11	99,07	87,1	92,11	83,75	79,79	889,05
E4	78,16	91,78	91,91	106,16	60,16	107,25	87,87	96,64	85,3	79,65	881,88
E5	70,78	81,87	90,87	108,42	64,45	79,85	82,65	86,13	87,31	83,82	836,14



**Figura 6:** Comportamento da previsão das 2 ações cujo melhor resultado foi usando E3 no ano 2011: ENBR3 e HYPE3. A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão

parte inferior do intervalo previsto então recomenda-se comprar ações. A Figura 6 mostra a previsão para o ano 2011 de ENBR3 e de HYPE3. A Fig. 6 mostra que a ação ENBR3 neste ano, teve muitas oscilações sem uma tendência definida de descida ou de subida, estas oscilações são de maior profundidade que as oscilações observadas na Fig. 5. Nos episódios de “escadinha” não consegue definir uma tendência. No caso da ação HYPE3, Fig. 6, embora apresente um comportamento claro de descida, os degraus da “escadinha” são grossos, portanto, as estratégias E4 e E5 são menos proveitosas. Além disso, observa-se na Fig. 6, que na maioria dos casos, tanto ENB3 quanto HYPE3, os trechos de intervalos iguais iniciam em preço baixo, sobe e baixa de preço, fazendo com que a previsão especulativa da E3 esteja correta para compra.

Da Tabela 4 observa-se que BBAS3, CYRE3, PSSA3, DTEX3 e ITUB4 tiveram melhor comportamento usando a E4. Como foi descrito anteriormente E4, diferencia-se de E2 e E3, porque quando o preço atual pertence ao intervalo previsto, compara-se o intervalo atual com o intervalo previsto, tentando aproveitar-se do comportamento em “escadinha” fornecido pelo método. Na E4 quando o intervalo atual e o previsto são iguais, a estratégia não faz nada.

A Fig. 7 mostra a previsão para o ano 2011 de BBAS3, CYRE3, PSSA3, DTEX3 e ITUB4. Na Fig. 7, ação BBAS3, observa-se tendências claras de descida detectadas pelo método de previsão através de mudança nos intervalos atual e futuro, forma de “escadinha” com degraus mais finos. Também, observa-se que nos trechos iguais a ação sobe, desce e sobe o que faz com que a previsão especulativa de E3 e E5 não seja vantajosa.

A ação CYRE3, Fig. 7, tem comportamento claro de des-

cida detectado pelo método de previsão, forma de “escadinha” com degraus mais finos, e tem um trecho longo de intervalos constantes e também outros trechos, cujo comportamento não define uma tendência e é aleatório. Portanto a previsão especulativa fica com um comportamento pior.

Na Fig. 7, mostra-se PSSA3, a qual apresenta comportamento claro de descida e de subida, vários destes movimentos são detectados pelo método de previsão, forma de “escadinha” com degraus mais finos. Os trechos iguais não tem um comportamento definido fazendo com que a previsão especulativa seja errática. O comportamento de DTEX3, ilustrado na Fig. 7, tem um comportamento claro de descida detectado pelo método, forma de “escadinha” com degraus mais finos e nos intervalos iguais o comportamento da ação não apresenta um padrão portanto a ação especulativa de compra não é a acertada neste caso.

Com respeito a ITUB4, a figura mostra que esta ação tem comportamentos claros de descida detectados pelo método através de intervalos diferentes, forma de “escadinha” com degraus mais finos e também no momento de intervalos iguais não existe um comportamento padrão portanto apresenta melhor desempenho não realizar nenhuma ação especulativa.

Da Tabela 4 observa-se que PETR4 teve melhor comportamento usando a E5. Como foi descrito anteriormente E5 diferencia-se da E4, porque quando o intervalo atual e o previsto são iguais, então se tenho ações e as comprei por um valor mais baixo que o atual então recomenda-se vender as ações, do contrário, se tenho dinheiro e se o preço da ação hoje estiver na terceira parte inferior do intervalo previsto então recomenda-se comprar ações. A Fig. 8 mostra a previsão para o ano 2011 de PETR4. Comportamentos de descida são acompanhados pelo método de previsão, forma de “escadinha” com degraus mais finos e nos intervalos iguais, o comportamento da ação é de preço baixo, aumentar e baixar de novo. Portanto, é conveniente a ação especulativa encontrada na E5.

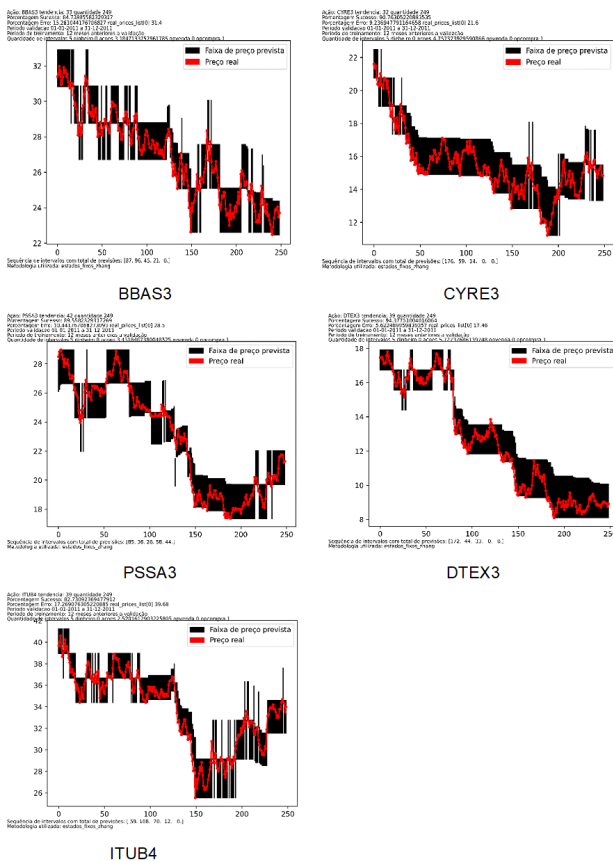
Dado que cada ano o comportamento será diferente para cada ação, Na Tabela 5 mostra-se os resultados de investimento levando em consideração o ano 2012. Neste ano, as E2 e E3 tiveram pior comportamento que a E1. Utilizando a melhor estratégia para cada ação então o resultado seria: R\$1.256,1.

Para confrontar o resultado da análise anterior, mostra-se na Fig. 9, DTEX3 e PETR4 no ano de 2012. DTEX3 em 2012, apresenta comportamento de “degraus” grossos e nos intervalos constantes não apresenta padrão por isso apresentou o melhor desempenho com a E2. De outra maneira, PETR4 em 2012, apresenta “degraus” finos e nos intervalos iguais não apresenta um padrão definido portanto o melhor comportamento será usando a estratégia



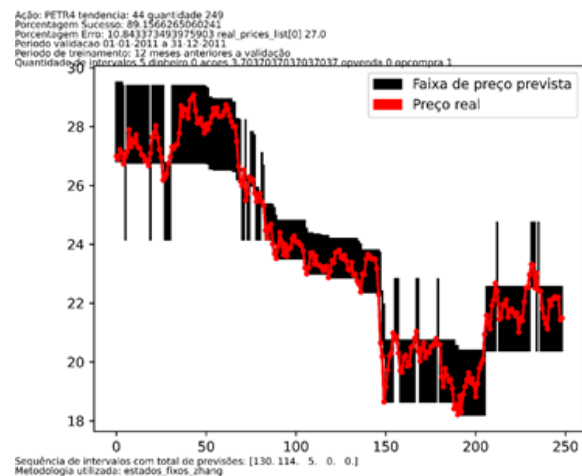
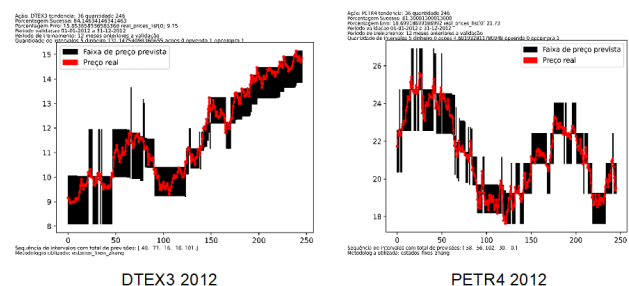
**Tabela 5:** Resultado de investimento de 5 estratégias diferentes, utilizando portfólio das ações BBAS3, CSAN3, CYRE3, ENBR3, HYPE3, PSSA3, DTEX3, ITUB4, PETR4 E VALE3 DTEX3 e ENBR3 no decorrer do ano 2012

Ação:	BBAS3	CSAN3	CYRE3	ENBR34	HYPE3	PSSA3	DTEX3	ITUB4	PETR4	VALE3	TOTAL
E1	105,24	145,15	115	41,98	183,82	108,96	157	98,72	91,16	100,76	1147,8
E2	105,24	104,7	118,98	100,6	156,2	108,96	128,5	98,72	91,15	100,76	1115,6
E3	102,04	113,41	126,67	108,68	163,6	103,34	116,6	99,96	89,08	99,08	1122,55
E4	99,18	118,18	181,7	104,45	168,33	122,65	123,33	98,03	92,66	93,42	1201,8
E5	108,04	124,93	188,69	37,55	173,08	127,84	123,83	101,35	86,2	102,28	1173,8

**Figura 7:** Comportamento da previsão das 5 ações cujo melhor resultado foi usando E4 no ano 2011: BBAS3, CYRE3, PSSA3, DTEX3 e ITUB4. A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão

E4.

Para poder aproveitar a análise feita, a heurística que pode ser adotada é apresentar as opções de todas as estratégias diariamente e analisar o gráfico parcial levando em consideração se a ação tem flutuações pequenas que o método não consegue detectar, se tem uma tendência de descida e subida que o método consiga detectar ou não e assim por diante. Se o gráfico parcial apresentar degraus “grossos” ou nenhum “degrau” significa que as estratégias apropriadas serão a E2 ou E3. Para fazer a opção por E2 ou E3 seria observar como está o comportamento da ação

**Figura 8:** Comportamento da previsão PETR4 cujo melhor resultado foi usando E5 no ano 2011. A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão**Figura 9:** Comportamento da previsão no ano 2012 para a ação DTEX3 cujo melhor resultado foi usando E2 e da ação PETR4 cujo melhor resultado foi usando E3. A curva vermelha ilustra os dados reais e as linhas verticais pretas correspondem aos intervalos com a maior probabilidade em cada previsão

nos intervalos iguais e escolher E3 caso o padrão de preço baixo, preço alto e preço baixo esteja se repetindo. Neste caso teríamos uma mistura ao longo do ano de E2 e E3. Caso o gráfico parcial apresente degraus “finos” quer dizer que as E4 e E5 são as mais apropriadas. Para selecionar

se será utilizado E4 ou E5 novamente tem que observar o comportamento dos intervalos iguais, escolhendo E5 caso o padrão de preço baixo, preço alto e preço baixo esteja se repetindo. Novamente neste caso teríamos uma mistura ao longo do ano de E4 e E5.

## 7 Conclusões

Neste trabalho foram implementados três métodos diferentes (intervalos fixos, intervalos variáveis, histograma e média móvel, intervalos variáveis e histograma) para fazer previsão de intervalos de preço de qualquer ação da bolsa de valores brasileira, permitindo escolher o período de tempo de construção da matriz de transição e o número de intervalos.

Na análise deste artigo foram escolhidas 10 ações da Bovespa e foram analisados 8 anos de comportamento no período de 2011 até 2019. A análise inicial constatou que, para este conjunto de dados, entre maior o período de tempo para construir a matriz de transição melhor foi a porcentagem de sucessos, neste caso o período foi fixado em 12 meses. Em seguida, foi verificado que entre menor número de estados ou intervalos usados maior a porcentagem de sucesso na previsão, neste caso de uso foi fixado em 5 intervalos ou estados. Em termos de porcentagem de sucessos, o método de intervalos fixos teve melhor comportamento com média total maior e desvio padrão menor, sendo que para os dados analisados foi obtido uma média de porcentagem de sucesso acima de 81% com desvio padrão de 4.8.

Foi mostrado que este método de previsão pode contribuir para melhorar o rendimento dos investimentos. Foram apresentadas 4 estratégias baseadas na informação do método desenvolvido. Dos resultados obtidos, recomenda-se aplicar uma estratégia diferente para cada ação do portfólio. Neste artigo mostra-se uma análise inicial onde são fornecidos subsídios para identificar qual seria a estratégia apropriada segundo o comportamento do método de previsão com respeito à ação específica.

Em resumo, o que é comum denominador dos artigos publicados na área é que não existe uma técnica que seja ótima para fazer previsão de preços nas ações. O que seria melhor opção é obter uma mistura de diversas técnicas, portanto, nos, autores deste artigo acreditamos que está abordagem de previsão de intervalos pode servir como complemento tanto a técnicas de cadeias de Markov que predizem tendências, tanto como para outras técnicas. Nosso interesse também vai orientado no sentido de complementar o modelo de previsão baseado em Cadeias de Markov Ocultas recentemente desenvolvido. Em pesquisas futuras, pretende-se propor uma técnica que envolva tanto cadeias de Markov discreta, cadeias de Markov ocultas e redes neurais.

## Acknowledgments

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

## Referências

- Abdullah, R. S. A. S. (2015). Applications of various stochastic models in financial prediction, *International Journal of Scientific and Innovative Mathematical Research (IJSIMR)* 3(3): 852–857.
- Adesokan, I. (2018). Markov chain asset pricing model for an emerging market. Disponível em <http://ir.jkuat.ac.ke/handle/123456789/4694>.
- Atsalakis, G. and Valavanis, K. (2009). Survey stock market forecasting techniques – part ii: Soft computing methods, *Expert Systems with Applications* 36: 5932–5941. <http://dx.doi.org/10.1016/j.eswa.2008.07.006>.
- Atsalakis, V. (2013). *Computation Optimization in Economics and Finance Research Compendium*, 1 edn, Nova Science Publishers, Inc Editors, New York.
- Bairagi, A. and Kakaty, S. (2015). Analysis of stock market price behaviour: a markov chain approach, *International Journal of Recent Scientific Research* 6(10): 7061–7066. Disponível em <https://recentscientific.com/sites/default/files/3581.pdf>.
- Bhusal, M. K. (2017). Application of markov chain model in the stock market trend analysis of nepal, *International Journal of Scientific & Engineering Research* 8(10).
- Choji, D., Eduno, S. and Kassem, G. (2013). Markov chain model application on share price movement in stock market, *Computer Engineering and Intelligent Systems* 4(10): 84–95. Disponível em <https://iiste.org/Journals/index.php/CEIS/article/view/7972/8139>.
- Dametto, R. (2018). Estudo da aplicação de redes neurais artificiais para predição de séries temporais financeiras. Disponível em <https://repositorio.unesp.br/handle/11449/157058>.
- de Oliveira, F. A., Nobre, C. N. and Zárata, L. E. (2013). 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* 40(18): 7596–7606. <https://doi.org/10.1016/j.eswa.2013.06.071>.
- Esbond, G. and Saporu, F. (2017). Introducing the use of markov chain model for informed decision making for investors in a stock market environment, *International Journal of Science and Research* 6(8): 1013–1019. Disponível em <https://www.ijsr.net/archive/v6i8/ART20171182.pdf>.
- Fritiyanto, A. and Lestari, T. (2018). Application of markov chain to stock trend: A study of pthm sampoerna, tbk, *Proceedings of the 3rd Annual Applied Science and Engineering Conference, IEMSS, Orlando, Florida, USA*. <http://dx.doi.org/10.1088/1757-899X/434/1/012007>.
- Kulkarni, V. (1999). *Modeling and Analysis, Design, and Control of Stochastic Systems*, 1 edn, Springer-Verlag, New York.

- Mitra, S. and Riggieri, M. (2011). Predicting stock prices, *Qualifying Project Report for Degree of Bachelor of Science MQP HS1 3699*, Worcester Polytechnic Institute, Worcester, Massachusetts.
- Neto, W. C. B., de Aguiar Salvi, A. and de Souza, W. P. (2020). Hybrid neural networks applied to brazilian stock market, *Revista de Informática Teórica e Aplicada* 7(2): 42–65. <https://doi.org/10.22456/2175-2745.88911>.
- Otieno, S., Otumba, E. and Nyabwanga, R. (2015). Application of markov chain to model and forecast stock market trend: a study of safaricom in nairobi securities exchange, kenya, *International Journal of Current Research* 7(4): 14712–14721. Disponível em <http://www.journalcra.com/sites/default/files/issue-pdf/8260.pdf>.
- Padhy, R., Patra, M. and Satapathy, S. (2011). Rdbms to nosql: Reviewing some next-generation non-relational databases, *International Journal of Advanced Engineering Science and Technologies* 11(1): 15–30. Disponível em <http://www.ijaest.iserp.org>.
- Redzwan, N., Musa, N., Latif, Y. A., Latip, A. H. A. and Rahman, I. N. A. (2019). Stock market analysis during election period in malaysia, *International Journal of Business and Economy* 1(2).
- Sezer, O. B., Gudelek, M. U. and Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, *Applied Soft Computing* 90: 106181. <https://doi.org/10.1016/j.asoc.2020.106181>.
- Sukkati, R. and Torse, D. A. (2019). Stock market forecasting techniques: A survey, *International Research Journal of Engineering and Technology (IRJET)* 6(5). Disponível em <https://www.irjet.net/archives/V6/i5/IRJET-V6I5577.pdf>.
- Sukkatti, R. and Torse, D. A. (2019). Stock market forecasting techniques: A survey, *International Research Journal of Engineering and Technology* 6(5). Disponível em <http://www.jatit.org/volumes/Vol46No1/4Vol46No1.pdf>.
- Sultan, Q., Fatima, K. and Ahmed, J. (2019). Application of markov 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"* XL(1): 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%27s\\_Stock\\_Market](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%27s_Stock_Market).
- Sun., Y. (2020). Index forecast study based on amended weighted markov chain in china, *International Journal of Trade, Economics and Finance* 11(5). <https://doi.org/10.18178/ijtef.2020.11.5.674>.
- Troiano, L. and Kriplani, P. (2010). Predicting trend in the next-day market by hierarchical hidden markov model, pp. 199–204. <https://doi.org/10.1109/CISIM.2010.5643663>.
- Vasanthi, S., Subha, M. and Nambi, T. (2011). An empirical study on stock index trend prediction using markov chain analysis, *Journal of Banking Financial Services and Insurance Research* 1(1): 72–87. Disponível em <https://www.indianjournals.com/ijor.aspx?target=ijor:jbfisr&volume=1&issue=1&article=005>.
- Zhang, D. and Zhang, X. (2009). Study on forecasting the stock market trend based on stochastic analysis method, *International journal of business management* 4(6): 163–160. Disponível em <https://www.ccsenet.org/journal/index.php/ijbm/article/view/2336>.