

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CURSO DE ENGENHARIA DA COMPUTAÇÃO

RICARDO JUNIOR FIORAVANTE

**UMA COMPARAÇÃO DE MÉTODOS ESTATÍSTICOS E DE
APRENDIZADO DE MÁQUINA NA PREDIÇÃO DA COTAÇÃO DA
BITCOIN**

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO
2021

RICARDO JUNIOR FIORAVANTE

**UMA COMPARAÇÃO DE MÉTODOS ESTATÍSTICOS E DE
APRENDIZADO DE MÁQUINA NA PREDIÇÃO DA COTAÇÃO DA
BITCOIN**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Da Computação da Universidade Tecnológica Federal do Paraná - UTFPR Campus Toledo, como requisito parcial para a obtenção do título de Bacharel em Engenharia Da Computação.

Orientador: Prof. Dr. Gustavo Henrique Paetzold
Universidade Tecnológica Federal do Paraná

**TOLEDO
2021**



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Toledo
Coordenação do Curso de Engenharia Eletrônica



TERMO DE APROVAÇÃO

Título do Trabalho de Conclusão de Curso Nº 70

**UMA COMPARAÇÃO DE MÉTODOS ESTATÍSTICOS E DE
APRENDIZADO DE MÁQUINA NA PREDIÇÃO DA COTAÇÃO DA
BITCOIN**

por

Ricardo Junior Fioravante

Esse Trabalho de Conclusão de Curso foi apresentado às **14h00 do dia 2 de Dezembro de 2021** como **requisito parcial** para a obtenção do título de **Bacharel em Engenharia Da Computação**. Após deliberação da Banca Examinadora, composta pelos professores abaixo assinados, o trabalho foi considerado **APROVADO**.

Prof. Dr. Daniel Cavalcanti Jeronymo
UTFPR

Prof. Dr. Gustavo Henrique Paetzold
UTFPR

Prof. Dr. Elder Elisandro Schemberger
UTFPR

O termo de aprovação assinado encontra-se na coordenação do curso

Toledo, 2 de Dezembro de 2021

AGRADECIMENTOS

Aos meus pais Ricardo e Vanice, e meu irmão Vinicius, pelo apoio incondicional e toda tranquilidade, união e conforto que me proporcionaram durante minha trajetória. A força e amor de vocês foi e é meu combustível para seguir em frente.

Ao meu orientador, Prof. Dr. Dr. Gustavo Henrique Paetzold, por ser uma inspiração enquanto pessoa e professor, por toda a paciência e dedicação, pelos ensinamentos técnicos e de vida, por todos os conselhos e compartilhamento de sabedoria, tempo e experiência, desenvolvendo pesquisas e trabalhos.

A todos os meus amigos e colegas, próximos ou nem tanto, que de alguma forma me ensinaram e compartilharam comigo sentimentos e vivências. Em especial, menciono meus grandes amigos Lucas Mendes e Luís Moromisato. Como também Eduardo Weber, Matheus Lino, Thales Ryu Ito, Hugo Leuchs, Jeison Sanches, Ronaldo Schlogl, Willian Rodrigo Huber, Felipe Stark, Fellipe Reis, Marcos Aurélio, Rafael Albarello, Jordano Lahn, Abel Menezes, Thais Mayara e Thiago Vinney.

Por fim, à Universidade Tecnológica Federal do Paraná, por ser palco de uma das experiências mais desafiadoras que tive, por todo conhecimento e sobretudo autoconhecimento, me apresentando forças que não imaginaria que tenho. Todos os professores e colaboradores, que em conjunto tornaram essa experiência única. Uma menção especial ao professor Cassius Rossi De Aguiar, que me ensinou muito e me tornou resiliente.

A todos e todas que aqui mencionei e também a quem não tive a oportunidade de mencionar, deixo a minha eterna gratidão.

Eu não quero acreditar, eu quero saber. (Carl Sagan).

RESUMO

O Bitcoin, lançado em 2009, foi a primeira criptomoeda e atualmente conta com mais de 50% de capitalização. Consequentemente vem chamando atenção nos últimos anos devido a sua alta valorização nos últimos anos chegando a valer \$66,000 USD. Porém, devido a sua alta volatilidade, seu valor é difícil de prever e, por conta disso não existem muitos registros para tentar prever seu valor futuro computacionalmente. A previsão do preço do Bitcoin é um desafio tanto para teoria econômica quanto computacional, porém, uma estratégia confiável de previsão de seu valor futuro poderia ter grande aplicabilidade em diversas áreas da indústria, como por exemplo plataformas de investimento e análise de mercado. Neste trabalho busca-se comparar sistemas computacionais capazes de prever valores futuros do Bitcoin por meio de algoritmos estatísticos e de aprendizado de máquina com exatidão e confiabilidade. Para o desenvolvimento do projeto foram utilizados algoritmos como: auto-regressivo (em inglês, AR), médias móveis (em inglês, MA), auto-regressivo integrado de médias móveis (em inglês, ARIMA), suavização exponencial simples (em inglês, SES), Prophet e rede de memória a longo prazo (em inglês, LSTM), para efeito de comparação. A previsão se deu pelo preço de fechamento, diário e de uma hora, em dólar. O banco de dados para essas variáveis foram adquiridos pelo site Kaggle. Para treinamento serão utilizados 80% dos dados e para teste 20% e por último foram feitos, uma análise de desempenho e alguns sistemas de compra e venda para comparação dos algoritmos. Assim, espera-se beneficiar investidores ao diminuir a incerteza na hora de tomar decisões.

Palavras-chave: Criptomoeda. Redes Neurais Recorrentes. Regressão. Séries Temporais.

ABSTRACT

Bitcoin, launched in 2009, was the first cryptocurrency and currently has more than 50 % capitalization. Consequently, it has been attracting attention in recent years due to its high appreciation in recent years reaching \$66,000 USD. However, due to its high volatility, its value is difficult to predict and, because of that, there are not many records to try to predict its future value computationally. Bitcoin price prediction is a challenge for both economic and computational theory, however, a reliable strategy for predicting its future value could have great applicability in several areas of the industry, such as investment platforms and market analysis. This work aims to compare computer systems capable of predicting future Bitcoin values through statistical and machine learning algorithms with accuracy and reliability. Algorithms such as: autoregressive (AR), moving averages (MA), integrated autoregressive of moving averages (ARIMA), simple exponential smoothing (SES), Prophet and long-term memory network (LSTM) were used to develop the project for comparison. The prediction was made by the closing price, daily and hourly, in dollars. The database for these variables were acquired from the Kaggle website. For training 80% of the data was used and for testing 20%, and finally, a performance analysis and some systems of buying and selling were made to compare the algorithms. Thus, beneficiaries are expected to reduce uncertainty when making decisions.

Palavras-chave: Cryptocurrency. Recurring Neural Networks. Regression. Time Series.

LISTA DE FIGURAS

Figura 1 – Cotação do Bitcoin em USD, nos últimos 7 anos.	1
Figura 2 – Estrutura interna da <i>Blockchain</i>	6
Figura 3 – Índice S&P 500.	8
Figura 4 – Estrutura interna do LSTM.	14
Figura 5 – Diagrama das etapas do projeto.	17
Figura 6 – Decomposição.	18
Figura 7 – Autocorrelação.	19
Figura 8 – Separação dos dados. Em vermelho o treino e em azul o teste.	20
Figura 9 – Média e desvio padrão do preço sem nenhuma diferenciação	20
Figura 10 – Média e desvio padrão do preço com uma diferenciação	21

LISTA DE TABELAS

Tabela 1 – Matriz de confusão	16
Tabela 2 – Descrição Das Variáveis	18
Tabela 3 – Métricas Para Treino - Period. 1D	23
Tabela 4 – Métricas Para Previsão - Period. 1D	23
Tabela 5 – Métricas Para Treino - Period. 1H	24
Tabela 6 – Métricas Para Previsão - Period. 1H	24
Tabela 7 – Melhores Parâmetros Encontrados - Period. 1D	24
Tabela 8 – Melhores Parâmetros Encontrados - Period. 1H	25
Tabela 9 – Desempenho do Sistema Direcional para cada modelo testado	26
Tabela 10 – Detalhes Do Sistema Direcional	26
Tabela 11 – Desempenho do Sistema de Votação para cada modelo testado	27
Tabela 12 – Detalhes Do Sistema de Votação	27
Tabela 13 – Desempenho do Sistema de <i>Buy and hold</i> para cada modelo testado	27

LISTA DE ABREVIATURAS E SIGLAS

AR	<i>Autoregressive</i>
MA	<i>Moving Average</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
LSTM	<i>Long short-term memory</i>
SES	<i>Simple Exponential Smoothing</i>
1D	1 Dia
1H	1 Hora

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	3
2.1	OBJETIVO GERAL	3
2.2	OBJETIVOS ESPECÍFICOS	3
3	JUSTIFICATIVA	4
4	REVISÃO DA LITERATURA	5
4.1	CRIPTOMOEDAS	5
4.1.1	BITCOIN	5
4.1.2	BLOCKCHAIN	6
4.2	ESTADO DA ARTE	7
4.3	SÉRIES TEMPORAIS	7
4.3.1	DECOMPOSIÇÃO	8
4.3.2	AUTOCORRELAÇÃO	9
4.3.3	ESTACIONARIEDADE E DIFERENCIAÇÃO	9
4.3.4	RAIZ DE TESTE UNITÁRIO	10
4.4	TÉCNICAS ESTATÍSTICAS	10
4.4.1	AUTO-REGRESSÃO	10
4.4.2	MÉDIA MÓVEL	11
4.4.3	AUTO-REGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS	11
4.4.4	SUAVIZAÇÃO EXPONENCIAL SIMPLES	11
4.4.5	PROPHET	12
4.5	TÉCNICAS DE APRENDIZADO DE MÁQUINA	13
4.5.1	LSTM	13
4.6	MEDIDAS DE DESEMPENHO	15
4.6.1	MATRIZ DE CONFUSÃO	16
5	MATERIAIS E MÉTODOS	17
5.1	COLETA DE DADOS	18
5.2	EXPLORAÇÃO DOS DADOS	18
5.3	PREPARAÇÃO DOS DADOS	19
5.4	CRIAÇÃO DE MODELOS	21
5.5	ANÁLISE INTRÍNSECA	22
5.6	ANÁLISE EXTRÍNSECA	22

6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	23
6.1	ANÁLISE INTRÍNSECA	23
6.2	ANÁLISE EXTRÍNSECA	25
7	CONCLUSÃO	29
	Referências	30

1 INTRODUÇÃO

A Bitcoin é um sistema de pagamento eletrônico *Peer-to-Peer* (Ponto a ponto), proposto em 2008, pelo pseudônimo de Satoshi Nakamoto, sendo que a identidade real de quem concebeu é desconhecida. Seu sistema é baseado em criptografia, permitindo que duas partes façam uma transação sem a necessidade de uma terceira parte para a validação sendo capaz de resolver o problema do gasto duplo, em que um usuário consegue gastar as mesmas moedas digitais mais de uma vez. Além disso, foi a primeira criptomoeda a utilizar, uma tecnologia descentralizada com blocos encadeados e seguros que guardam algum conteúdo, a *blockchain* (NAKAMOTO, 2008).

Atualmente existem varias criptomoedas e muitas delas ainda utilizam a tecnologia *blockchain*. No entanto, a Bitcoin ainda é a mais popular entre elas, com um mercado de capitalização de 1.1 trilhões de dólares e ocupando uma dominância de 43% do mercado de criptomoedas (COINMARKETCAP, 2021). Essa popularização se deve ao fato da sua tecnologia inovadora, a *blockchain*, e a sua valorização nos últimos anos que chegou a valer \$ 66.000 USD, como mostra a Figura 1. Essa valorização atraiu muitas pessoas que viram uma potencial forma de investimento (COINMARKETCAP, 2021).

Figura 1 – Cotação do Bitcoin em USD, nos últimos 7 anos.



Fonte: COINMARKETCAP (2021)

Muitos estudos foram conduzidos para identificar os principais fatores que determinam o preço da Bitcoin, (MCNALLY; ROCHE; CATON, 2018), (ROY; NANJIBA; CHAKRABARTY, 2018). Entretanto, poucas pesquisas foram feitas para tentar estimar ou prever o seu preço, o que somado a sua alta volatilidade e a períodos de bolha especulativa em que o ativo se desvia fortemente do valor intrínseco correspondente desse mesmo, a tornam em um investimento de alto risco (KALYVAS et al., 2020). Estudos anteriores tentaram prever o preço da Bitcoin com algoritmos de aprendizado de máquina e estatísticos, indicando assim que esses algoritmos possam ser relevantes como solução. Por exemplo temos MCNALLY, ROCHE e CATON (2018), que usaram apenas os modelos RNN (*Recurrent Neural Network*), LSTM e ARIMA, e obtiveram respectivamente uma acurácia (proximidade entre o valor obtido e o valor verdadeiro)

na classificação de 50.25%, 52% e 50.05%, com periodicidade de 1 dia. Desse modo fica evidente que modelos como RNN e LSTM são eficientes, porém com resultados não muito impressionantes devido a alta volatilidade na tarefa (MCNALLY; ROCHE; CATON, 2018).

Este trabalho tem o objetivo de comparar a eficácia de diversos métodos estatísticos e de aprendizado de máquina na predição do valor da criptomoeda Bitcoin. Para isso alguns algoritmos como: AR, MA, ARIMA, SES, Prophet e LSTM serão desenvolvidos e comparados na tentativa de abordar tal problema. Como variável de entrada somente o preço de fechamento será usado, pois esse é o interesse da maioria dos investidores, tanto para os modelos com periodicidade de 1 dia como de 1 hora. Já para o treino serão separados dados do período de 08/10/2015 até 15/05/2019 (80%) e para testes do período de 16/05/2019 até 09/04/2020 (20%). Ao final, para quantificar a usabilidade do modelo, foram também implementados sistemas que usam os algoritmos testados para realizar a compra e venda do Bitcoin na prática.

Este trabalho está organizado da seguinte forma: os Capítulos 2 e 3, estabelecem os objetivos geral e específicos, respectivamente. No Capítulo 4, encontra-se a revisão da literatura existente, e no Capítulo 5 são abordados os materiais e métodos que foram utilizados no desenvolvimento deste trabalho, abrangendo aspectos de coleta e preparação dos dados, treino para parametrização dos modelos, análise intrínseca e extrínseca e obtenção dos resultados. No Capítulo 6 são mostrados as análises e os resultados para o trabalho e conclusões são expostas no Capítulo 7. Por fim, são apresentadas as referências usadas ao longo deste trabalho

2 OBJETIVOS

2.1 OBJETIVO GERAL

Comparar o desempenho de diversos algoritmos estatísticos e de aprendizado de máquina na predição da cotação da Bitcoin, tanto de forma intrínseca, quanto extrínseca.

2.2 OBJETIVOS ESPECÍFICOS

- Implementar modelos de regressão (AR, MA, ARIMA, SES e Prophet) para prever o preço da Bitcoin.
- Implementar um modelo de aprendizado de máquina (LSTM) para prever o preço da Bitcoin.
- Análise intrínseca dos modelos, por meio de métricas de desempenho para regressão.
- Análise extrínseca dos modelos, por meio de sistemas de compra e venda
- Conceber um sistema completo de compra e venda da Bitcoin usando os melhores modelos e políticas de compra e venda disponíveis dentre as testadas

3 JUSTIFICATIVA

Hoje existem vários sistemas automáticos e semiautomáticos de investimento, como ZuluTrade ((ZULUTRADE, 2021)) e MetaTrader 4 ((METATRADER4, 2021)), é um mercado enorme que visa integrar vários tipos de produtos (de renda fixa a ações de empresas). No entanto, criptomoedas são difíceis de incorporar nessas carteiras de investimento por terem sido pouco estudadas ainda ou por falta de regulamentação. Portanto, a criação de modelos que incorporam sistemas eficazes de predição das mesmas poderiam aumentar o potencial de diversificação oferecido por estas plataformas de investimento.

O desenvolvimento de tal sistema é desafiante tanto para teoria econômica quanto para a computacional, uma vez que oferece a oportunidade da descoberta de novas aplicações, metodologias, análises e utilidades. Assim com a resolução do problema, implementando um modelo preditivo, espera-se beneficiar investidores ao diminuir a incerteza na hora de tomar decisões.

4 REVISÃO DA LITERATURA

O capítulo a seguir tem como objetivo apresentar, de maneira objetiva, alguns conceitos básicos para o entendimento desse trabalho.

4.1 CRIPTOMOEDAS

Desde o lançamento da Bitcoin em 2009 o mercado de criptomoedas tem crescido exponencialmente, atingindo um pico de 800 bilhões de capitalização de mercado em 2018 (COINMARKETCAP, 2021). O que começou como um meio de pagamento descentralizado de pessoa pra pessoa para alguns, terminou como um tópico com crescente atenção da mídia e uma oportunidade de investimento para muitos. Esse grande crescimento atraiu também a atenção de autoridades reguladoras que buscam principalmente leis de taxação e contra lavagem de dinheiro, além de investidores e recentemente de grandes corporações que estão entrando nesse mercado. Investidores individuais geralmente são atraídos pelas histórias de sucessos de investidores antigos através da mídia, a qual parcialmente tem contribuído para direcionar o preço da Bitcoin. Agentes econômicos e fundos também vem buscando esse mercado como maneira de diversificar e se assegurar, assim também como negociadores (*traders*) que se beneficiam da alta volatilidade (PAPADIMITRIOU; GOGAS; GKATZOGLOU, 2020).

4.1.1 BITCOIN

Toda tecnologia e criptografia por trás da Bitcoin já o faz diferente das moedas fiduciárias comuns, principalmente por causa da sua representação em possuir valor digital e sua descentralização virtual. Nesse sentido, não há consenso entre os acadêmicos em usar o termo moeda, para se referir ao Bitcoin. Alguns aspectos da Bitcoin diferem das moedas fiduciárias tradicionais e devem ser analisados como: meio de troca, unidade de conta e reserva de valor (NEVES, 2020).

O uso do dinheiro virtual tem aumentado como meio de troca, permitindo compra e venda em troca da moeda, nos ambientes de *e-commerce* onde grandes empresas como Microsoft e Subway tem oferecido como meio de pagamento nas compras online. A velocidade e o baixo custo das transações da Bitcoin, o anonimato da transferência e a transparência das transações guardadas na *blockchain* são aspectos muito positivos para promover a Bitcoin como moeda. Porém, há ainda aspectos negativos como falta de regulamentação, eliminando as possibilidades de acordos judiciais e dificultando o reembolso de operações prejudicadas contra criptomoedas (NEVES, 2020).

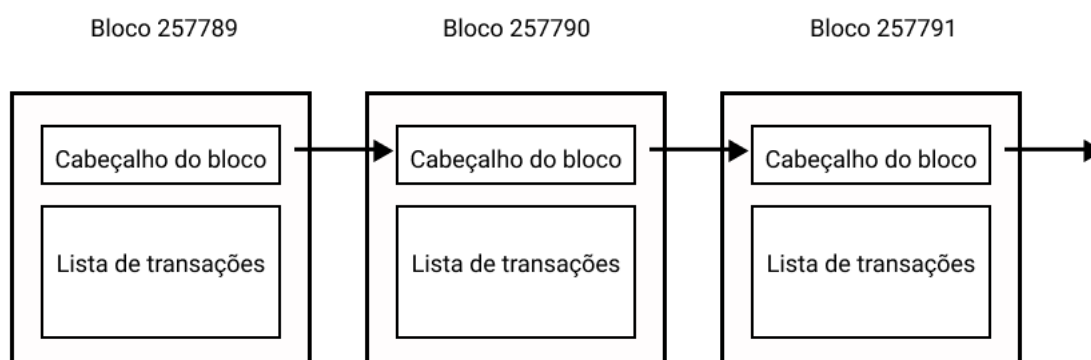
A respeito da unidade de conta, onde bens e serviços são precificados, a alta volatilidade do preço da Bitcoin é muito custosa para preços denominados em suas unidades monetárias. Por sua vez, essa é a maior diferença se tratando de moedas fiduciárias e a Bitcoin. Já se tratando da

reserva de valor, característica essa de manter o valor de compra ao longo do tempo, a Bitcoin tem duas vantagens importantes sobre outras moedas: o fato da sua oferta ser predeterminada pela sua plataforma e ser protegida, uma vez que todas as transações feitas pela blockchain são quase impossíveis de serem alteradas. E novamente sua volatilidade é uma desvantagem nesse ponto, assim como possíveis ataques a rede (NEVES, 2020).

4.1.2 BLOCKCHAIN

A Bitcoin foi a primeira cripto moeda a usar a tecnologia *blockchain*. Ela é composta por uma corrente de blocos que usa assinaturas digitais, onde cada proprietário transfere seus Bitcoins para outro usando sua assinatura digital, que contém informações sobre o *hash* da transação e a chave pública do próximo proprietário, adicionando assim essa informação no final do bloco. A rede do Bitcoin é composta por vários computadores que são conectados na internet e para evitar que uma terceira parte valide suas transações, foi implementado um mecanismo de "prova de trabalho", onde os nós tentam resolver cálculos matemáticos para gerar outro bloco na cadeia e validar as transações. O bloco então não pode ser refeito sem que todo o cálculo seja refeito, além de gerar uma recompensa a quem conseguiu resolver o cálculo. Por fim a rede possui um limite de 21 milhões de Bitcoins, para evitar a geração de inflação (NAKAMOTO, 2008).

Figura 2 – Estrutura interna da *Blockchain*.



Fonte: Autoria própria.

A estrutura *Blockchain* da Bitcoin é composta por blocos ligados uns aos outros em sequência. Cada bloco possui uma lista de transações e um cabeçalho como visto na Figura 2. Dentro desse cabeçalho possui um campo chamado "*Nonce*", em que é variado o seu valor para ser aplicado a uma função de *hash* (criptografia) ao seu cabeçalho até que o resultado seja composto pela quantidade de zeros iniciais necessárias para a "prova de trabalho". Dessa maneira fica fácil de verificar se a prova esta correta e impede que os blocos anteriores sejam alterados sem que todos eles sejam refeitos, já que no cabeçalho possui também o *hash* do bloco anterior.

4.2 ESTADO DA ARTE

Relativamente poucos estudos foram conduzidos para tentar prever os preços do Bitcoin. Yenidogan et al. (2018) conduziram um estudo comparativo entre os métodos ARIMA e PROPHET e mostraram que o último possui desempenho superior, apresentando um R^2 (coeficiente de determinação, é a medida de ajuste de um modelo estatístico linear generalizado) de 94,5% em relação a 68,1% do ARIMA. O estudo também analisou a correlação do Bitcoin com outras moedas fiduciárias como Euro, libra e Yen, e depois de uma análise de correlação optou por usar o Yen como regressor adicional para o modelo, para o treino de ambos os modelos foram utilizadas maneiras automáticas dadas pelas bibliotecas escolhidas. Phaladisailoed e Numnonda (2018) comparam alguns modelos de regressão como *Huber* e *Theil-Sen*, e com alguns de aprendizado de máquina, como GRU (*Gated Recurrent Unit*) e LSTM. Os resultados mostraram um erro quadrático médio tão baixo quanto 0,00002 e um R^2 tão alto quanto 99,2% dado pelo GRU. Para esse estudo, o tempo gráfico de 1 minuto foi escolhido e observou-se que talvez somente o preço não seja o suficiente pois, as reações das mídias sociais, políticas e leis que cada país anuncia pode influenciar na movimentação dos preços.

Liantoni e Agusti (2020) levaram em conta um modelo de suavização dupla (em inglês, DES) e nele usaram o erro percentual médio absoluto como métrica avaliativa, obtendo um resultado de 2.89% para o treino com um alfa (parâmetro de suavização) de 0.9, e 0.0373% na previsão do preço de fechamento na data 01/01/2020. Roy, Nanjiba e Chakrabarty (2018) usaram o ARIMA e também usaram seus modelos derivados como o MA e AR, fazendo uma análise da série temporal e definindo assim os parâmetros para os modelos propostos. O resultado da acurácia obtida da previsão de 10 dias consecutivos foi de 87.58%, 89.25% e 90.31% para o MA, AR e ARIMA respectivamente, sendo o ARIMA o que mostrou melhores resultados.

4.3 SÉRIES TEMPORAIS

As séries temporais são séries de observações na ordem do tempo, onde geralmente tem-se uma mesma periodicidade. Um exemplo de séries temporais é mostrado na Figura 3. nela mostra o índice do S&P 500 (índice composto por quinhentos ativos cotados nas bolsas de NYSE ou NASDAQ), nos últimos meses. Para investigação podemos dividir ela em duas áreas, descritiva e preditiva. A descritiva ajuda a entender e explicar os padrões em que os dados se encontram. Já a preditiva, explica como serão os dados no futuro e, pode ser dividida ainda em 4 técnicas: Predição baseada em decomposição, suavização, regressão e aprendizado de máquina (KOTU; DESHPANDE, 2019).

Figura 3 – Índice S&P 500.



Fonte: (INVESTING, 2020)

4.3.1 DECOMPOSIÇÃO

A decomposição de uma série temporal é o processo de desconstrução em componentes, representando um fenômeno na série. Ela é composta dos seguintes componentes: tendência, sazonalidade e ruído. A técnica de decomposição se torna importante antes da predição, pois permite entender e descrever melhor os componentes que formam a série, para melhor adaptar o modelo. Uma descrição mais detalhada sobre cada componente pode ser observado logo abaixo (KOTU; DESHPANDE, 2019):

Tendência: Ela representa a mudança de um período ao outro e pode ser crescente ou decrescente. Por vezes, quando inserida num gráfico ela é simplesmente uma equação linear que mais se aproxima dos pontos.

Sazonalidade: É o comportamento ou padrão repetitivo dentro de uma série. Ele pode estar contido em horário, diário, semanal, mensal, anual e etc.

Ruído: Tudo que não for representado pela tendência ou sazonalidade é considerado ruído. Idealmente ele deve seguir uma distribuição normal e toda série possui ele.

A decomposição da série temporal pode ser classificada em aditiva ou multiplicativa, a depender de como os componentes se relacionam. No aditivo, os componentes são dispostos de uma maneira em que a soma dos componentes retorna a série temporal original, enquanto no

multiplicativo é a multiplicação desses componentes que produz esse retorno. Um indicativo entre a melhor escolha do tipo de decomposição pode ser observado pela mudança do nível da tendência ou da magnitude da sazonalidade ao longo do tempo. As fórmulas que descrevem os tipos de decomposição são descritas a seguir:

$$Y_t = T_t + S_t + E_t, \quad (1)$$

$$Y_t = T_t * S_t * E_t, \quad (2)$$

onde T_t , S_t e E_t são respectivamente tendência, sazonalidade e ruído. Y_t é a combinação desses componentes originando a série temporal original (KOTU; DESHPANDE, 2019).

4.3.2 AUTOCORRELAÇÃO

Assim como a correlação mede a relação linear entre duas variáveis, a autocorrelação mede a relação linear entre valores atrasados da série temporal. Com ela podemos notar algumas características na série temporal. Quando temos tendência, por exemplo é comum observar um decaimento desacelerado nos seus coeficientes. Seus coeficientes variam de -1 a 1 e ela pode ser representada através de um correlograma. A função pode ser definida como se segue:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}, \quad (3)$$

onde N é o tamanho da série temporal, k é a distância de atraso, x_t é o seu valor no tempo t , $\bar{x} = \sum_{t=1}^N \frac{x_t}{N}$ e r_k é o coeficiente de correlação entre as observações numa distancia k (CHATFIELD, 1996).

4.3.3 ESTACIONARIEDADE E DIFERENCIAÇÃO

Há situações em que se deseja que as propriedades estatísticas de uma série temporal sejam as mesmas ao longo da série, ou seja, queremos que ela tenha estacionariedade. Geralmente o termo estacionariedade é entendido como sendo de estacionariedade fraca ou também como de estacionariedade de covariância. Dessa maneira, o termo implica que a média e a autocovariância da série não muda através do tempo. Diferentes porções da série temporal, ditas estacionárias, podem então serem consideradas como tendo a mesma média, variância e covariância (WILKS, 2011). Um exemplo que atende aos requisitos de estacionariedade é o ruído branco (sinal aleatório), onde a série temporal possui condições de média igual a zero e variância constante. Nesse caso observa-se também que não há autocorrelação na série (HAMILTON, 1994).

Quando uma série temporal ou processo viola as características mencionadas anteriormente, de manter uma mesma média, variância e covariância, ela apresenta então a propriedade de ser não estacionária. Nesse sentido, tal particularidade é conhecida como processo integrado, onde um processo estocástico não estacionário para uma variável Y_t é chamado de processo

integrado de ordem d ou $I(d)$, se é preciso diferenciá-lo ($Y_t - Y_{t-1}$) ao menos d vezes para se tornar um processo estacionário. Ou seja, se Y_t for um processo não estacionário integrado de ordem 2, ou $I(2)$, significa dizer que teremos que diferenciar ao menos 2 vezes para obtermos um processo estacionário (RIBEIRO, 2014).

4.3.4 RAIZ DE TESTE UNITÁRIO

O teste de raiz unitária, *Augmented Dicky Fuller* (ADF), tenta descobrir o quão forte a série é determinada pela tendencia. Consideremos agora o processo AR(1) dado pela equação:

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t, \quad (4)$$

iremos testar se ϕ_1 é igual a 1. Subtraindo o termo Y_{t-1} de ambos os lados da equação, temos:

$$\Delta(Y_t) = \phi_1 Y_{t-1} - Y_{t-1} + \varepsilon_t = (\phi_1 - 1)Y_{t-1} + \varepsilon_t, \quad (5)$$

onde denotamos o termo $\phi_1 - 1$ como ψ . Podemos então testar se o termo ψ é igual a zero na Equação 6, pelo teste de *Augmented Dicky Fuller*. Onde o teste é calculado pela razão:

$$ADF = \frac{\psi}{\text{desviopadro}(\psi)}, \quad (6)$$

o qual o resultado é sempre um numero negativo (PROFILLIDIS; BOTZORIS, 2019).

Para a hipótese nula (H_0) do teste, temos que a série não é estacionaria, e para a hipótese alternativa (H_1), temos que a série é estacionaria. Uma interpretação do valor-P (em inglês, *P-value*) é dada a seguir:

- valor-P > 0.05: Aceita a hipótese nula
- valor-p < 0.05: Rejeita a hipótese nula

dessa maneira ao aceitar a hipótese nula, deve-se diferenciar a série para que ela se torne estacionaria (PROFILLIDIS; BOTZORIS, 2019).

4.4 TÉCNICAS ESTATÍSTICAS

Muitas técnicas estatísticas foram desenvolvidas para a previsão de séries temporais, algumas delas serão discutidas logo a seguir.

4.4.1 AUTO-REGRESSÃO

O AR é um modelo para séries temporais estacionárias que usa observações com um determinado atraso como entrada em uma equação de regressão para prever valores futuros. A equação que representa o modelo pode ser descrita como se segue:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad (7)$$

onde Y_t é o valor predito, c é uma constante, ϕ_p são os coeficientes encontrados na otimização através do treino do modelo, ε_t é o termo de erro e um ruído branco e p é a ordem do modelo, de maneira que Y_t representa $AR(p)$ (HAMILTON, 1994).

Ainda algumas restrições para o modelo se manter estacionário, para esses casos há limites nos valores a serem respeitados, como:

- Para um modelo $AR(1)$: $-1 < \phi_1 < 1$.
- Para um modelo $AR(2)$: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

Quando $p \geq 3$ as restrições são mais complicadas e geralmente a própria biblioteca que possui o modelo cuida de tais restrições (HYNDMAN; ATHANASOPOULOS, 2018).

4.4.2 MÉDIA MÓVEL

A MA é um modelo para séries temporais estacionárias que usa valores de erros de previsões passadas em uma equação de regressão como maneira de prever valores futuros. A equação que representa o modelo pode ser descrita como se segue:

$$Y_t = \mu + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-q}, \quad (8)$$

onde Y_t é o valor predito, μ é uma constante, ϕ_p são os coeficientes encontrados na otimização através do treino do modelo, ε_t é o termo de erro e um ruído branco e q é a ordem do modelo, de maneira que Y_t representa $MA(q)$ (HAMILTON, 1994).

4.4.3 AUTO-REGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS

O modelo ARIMA surge de uma extensão do modelo ARMA com a condição de ser diferenciado d vezes anteriormente para atender a condição de estacionariedade (MONTGOMERY CHERYL L. JENNINGS, 2008). O modelo ARMA é expresso como sendo uma combinação dos modelos AR e MA e é definido da seguinte maneira:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \mu + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-q}, \quad (9)$$

onde p é a ordem da parte auto-regressiva e q é a ordem da parte da média móvel, de maneira que Y_t representa $ARMA(p, q)$, e estendendo o modelo para ser diferenciado d vezes temos o $ARIMA(p, d, q)$ (HAMILTON, 1994).

O valor de diferenciação também tem efeito na predição dos intervalos. Quanto maior a ordem de diferenciação, mais rapidamente a ordem dos intervalos aumenta. Por exemplo, com $d = 0$, o desvio padrão da previsão se aproxima do desvio padrão histórico e a predição se torna a mesma (HYNDMAN; ATHANASOPOULOS, 2018).

4.4.4 SUAVIZAÇÃO EXPONENCIAL SIMPLES

A suavização exponencial simples leva em conta a previsão passada e ajusta ela com o seu erro. Ou seja, a previsão para o próximo período é dado por:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t), \quad (10)$$

onde \hat{y}_{t+1} é o valor previsto no tempo $t+1$, \hat{y} é o valor previsto no tempo t , y_t é o valor da série temporal no tempo t e α é coeficiente de suavização e varia de 0 á 1 (HYNDMAN ANNE B. KOEHLER, 2008).

Dessa maneira pode se notar que a previsão é simplesmente a previsão antiga mais um ajustamento do erro dela. Quando α é perto de 1, o erro incluso será maior e quando α é perto de zero o erro incluso é menor. A forma expandida do modelo é definida por:

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots, \quad (11)$$

onde \hat{y}_{t+1} representa uma média móvel ponderada de todos os valores passados com pesos decrescendo exponencialmente (HYNDMAN ANNE B. KOEHLER, 2008).

4.4.5 PROPHET

O modelo Prophet foi criado para lidar com problemas de negócios onde há séries temporais envolvidas, como predição ou análise, por exemplo (TAYLOR; LETHAM, 2017). Ele também foi desenvolvido para ter parâmetros intuitivos que podem ser ajustados sem necessariamente saber muito dos detalhes por baixo deles. O modelo é composto por 3 componentes que podem ser descritos como:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (12)$$

onde $y(t)$ é a série temporal, $g(t)$ é a função de tendência que modela mudanças não periódicas nos valores da série temporal, $s(t)$ representa mudanças periódicas como sazonalidade diária, semanal, anual entre outras, $h(t)$ representa o efeito de um feriado que ocorrem por vezes de maneira irregular e o ϵ_t representa mudanças que não são captadas pelo modelo (TAYLOR; LETHAM, 2017).

A função de tendência definida pelo modelo pode ser expressa como uma função de saturação dada da seguinte maneira:

$$\gamma_j = (s_j - m - \sum_{l < j} \gamma_l) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l}\right), \quad (13)$$

$$a_j(t) = \begin{cases} 1, & \text{se } t > s_j, \\ 0, & \text{caso contrário,} \end{cases} \quad (14)$$

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^\top \delta)(t - (m + a(t)^\top \gamma)))}, \quad (15)$$

onde $C(t)$ é o limite da curva em função do tempo; $k + a(t)^\top$ é a taxa de crescimento dado o tempo t ; δ é a mudança da taxa que o ocorre no tempo s_j , ela pode ser definida manualmente ou automaticamente pelo modelo; m é o parâmetro de *offset* e é ajustado pela função dada por γ_j (TAYLOR; LETHAM, 2017).

Para problemas que não exibem saturação de crescimento, e possuem crescimento contínuo a função é descrita da seguinte maneira:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma), \quad (16)$$

onde como antes k é a taxa de crescimento; δ é a taxa de ajustes; m é o parâmetro de *offset* e γ é dado por $-s_j \delta_j$ para manter a função contínua (TAYLOR; LETHAM, 2017).

A função de sazonalidade definida pelo modelo se baseia nas séries de Fourier expressa da seguinte maneira:

$$s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P})), \quad (17)$$

onde $s(t)$ é o valor da sazonalidade no tempo t , P representa o período da série, N é o tamanho da série temporal e a_n e b_n são coeficientes da série (TAYLOR; LETHAM, 2017).

Para representar a função que descreve os efeitos do feriado, consideramos uma matriz de regressões:

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)], \quad (18)$$

onde cada feriado i pertence a uma data D_i e é atribuído uma mudança k representado pela equação:

$$h(t) = Z(t)k, \quad (19)$$

para essa mudança ainda é incluso efeitos de janelas de dias em volta do feriado (TAYLOR; LETHAM, 2017).

4.5 TÉCNICAS DE APRENDIZADO DE MÁQUINA

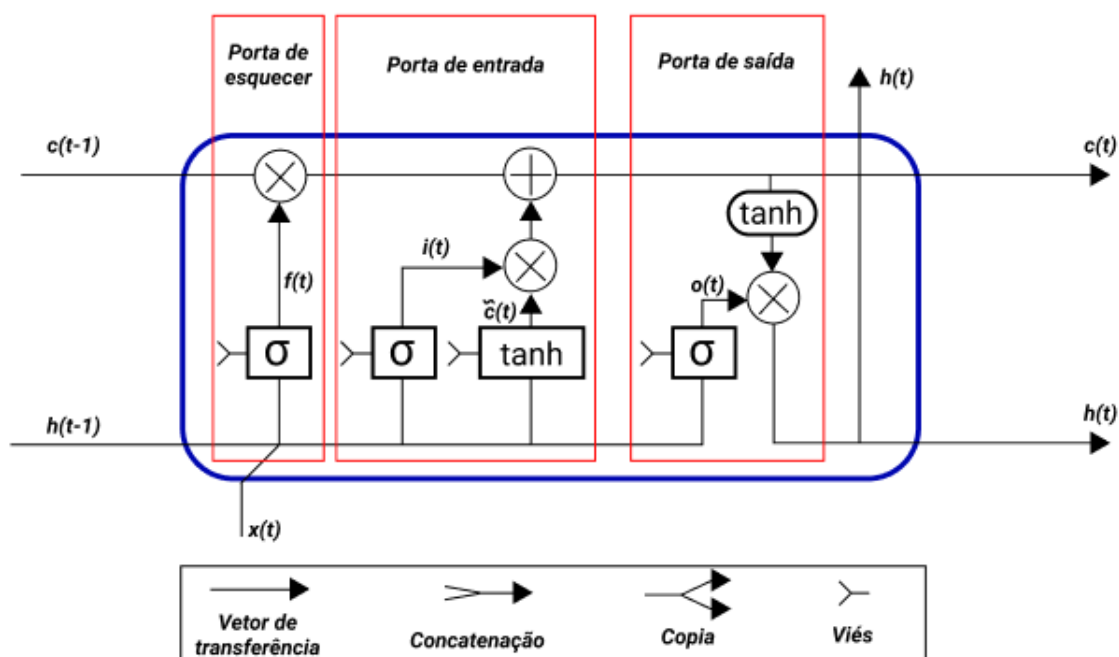
Aprendizado de máquina é uma sub área da inteligência artificial (IA) que consiste em sistemas que podem aprender com dados e identificar padrões e esse aprendizado pode ser dividido em supervisionado, não supervisionado e por reforço. O supervisionado é caracterizado pelo conjunto de dados ser totalmente rotulado e o modelo ser capaz de verificar se no treinamento a ação esta correta. Para o tipo não supervisionado o modelo deve aprender a encontrar e definir os rótulos e, por último, no aprendizado de reforço o modelo é recompensado por uma ação correta e punido por uma ação incorreta (MIRJALILI; FARIS; ALJARAH, 2020). Nesse trabalho será usado somente o supervisionado.

4.5.1 LSTM

O modelo *Long Short-Term Memory* (LSTM), o qual emprega uma arquitetura de rede neural recorrente (RNR), que consiste em nós interconectados adaptados para dados que envolvem sequencia, foi proposto originalmente por Hochreiter and Schmidhuber em 1997 e é amplamente usado nos dias atuais por ter uma superioridade em modelar, curtas e longas dependências nos dados (BIANCHI et al., 2017). O LSTM tenta resolver o problema da dissipação do

gradiente não impondo viés nas observações recentes, mas mantendo o fluxo de erro constante através do tempo. Ele se diferencia de outras arquiteturas de RNR apenas por possuir uma célula. Além disso, a versão mais simples do LSTM é composta também por uma porta de entrada, saída e de esquecer, a qual permite redefinir o seu estado. A célula lembra os valores de um intervalo arbitrário e as portas regulam o fluxo de informação associados com a célula (HOUDT; MOSQUERA; NÁPOLES, 2020).

Figura 4 – Estrutura interna do LSTM.



Fonte: Autoria Própria.

Em resumo, a arquitetura do LSTM consiste em um conjunto de blocos de memória ligados recursivamente. A ideia por trás do bloco de memória seria manter o estado durante um período e regular o fluxo de informação através das portas, de maneira que a saída do bloco seja recursivamente ligada a entrada do próximo bloco e suas portas (HOUDT; MOSQUERA; NÁPOLES, 2020). A Figura. 4 nos mostra essa arquitetura, que envolve o sinal de entrada $x(t)$, a informação corrente $h(t)$ e o de saída $y(t)$ no instante t da célula (YU et al., 2019).

O LSTM, como visto na figura.4, pode ser expresso como:

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f), \quad (20)$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i), \quad (21)$$

$$\tilde{c}_t = \tanh(W_{\tilde{c}_h} h_{t-1} + W_{\tilde{c}_x} x_t + b_{\tilde{c}}), \quad (22)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \quad (23)$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o), \quad (24)$$

$$h_t = o_t \cdot \tanh(c_t), \quad (25)$$

onde c_t é o estado da célula do LSTM; W_i , $W_{\tilde{c}}$, W_o são os pesos, b é o viés e o operador " \cdot " representa a multiplicação escalar de vetores. Ao atualizar o estado da célula, a porta de entrada pode decidir que informação nova pode ser guardada nela, e a porta de saída decidir que informação deve sair. A porta de saída pode decidir a informação que será retirada da célula. Quando o valor de f_t se aproximando de 1, ela guarda a informação e contrariamente quando se aproxima de 0 ela se livra dela (YU et al., 2019).

4.6 MEDIDAS DE DESEMPENHO

Para poder descrever o resultado do desempenho gerado pelas séries temporais algumas medidas são consideradas. Dessa maneira, definimos Y_t como uma observação no tempo t e F_t a previsão de Y_t . Então definimos o erro da previsão como $e = Y_t - F_t$. Há algumas formas de medidas que dependem da escala que são úteis quando comparadas com diferentes métodos para os mesmos dados, mas não devem ser usadas para comparação em diferentes conjuntos de dados (HYNDMAN; KOEHLER, 2006). Dentre essas medidas temos o:

$$\text{Raiz do Erro Quadrático Médio (RMSE em inglês)} = \sqrt{\frac{1}{t} \sum_{n=1}^t e_t^2}. \quad (26)$$

Há ainda algumas medidas que são baseadas nas porcentagens do erro, onde a porcentagem do erro é dado por $p_t = \frac{100e_t}{Y_t}$, que não dependem da escala e são frequentemente usados para comparar diferentes tipos de dados (HYNDMAN; KOEHLER, 2006). Uma delas é o:

$$\text{Erro Absoluto Médio Percentual (MAPE em inglês)} = \frac{1}{t} \sum_{n=1}^t |p_t|. \quad (27)$$

Temos também, a medida de variação dada pelo coeficiente de determinação R^2 , que representa a proporção da variabilidade dos valores previstos, explicada pelo modelo considerado no qual o valor de R^2 varia entre 0 e 1, e quanto mais perto de 1 melhor o ajuste do modelo considerado (SUBRAMANIAN; SILVA; COUTINHO, 2007). O coeficiente pode ser expresso da seguinte forma:

$$R^2 = \frac{\sum_{n=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{n=1}^N (y_i - \bar{y})^2}, \quad (28)$$

onde N é o tamanho da série, \hat{y}_i é o valor previsto, y_i é o valor real da série e \bar{y} é dado por $\frac{1}{N} \sum_{n=1}^N y_i$.

Por último, o critério de informação de Akaike (em inglês, AIC) usado para estimar a perda de informação de um modelo é dado por:

$$AIC = -2\ln(\hat{L}) + 2p, \quad (29)$$

onde \hat{L} , é o valor de máxima verossimilhança estimado pelo modelo e p é o numero total de parâmetros estimados para o modelo (MCELREATH, 2020)

4.6.1 MATRIZ DE CONFUSÃO

A matriz de confusão é dividida em 4 categorias de resultados, que são eles: Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN). Ela contém estatísticas sobre a real classificação e predição dos dados gerados pelos modelos (DENG et al., 2016).

Tabela 1 – Matriz de confusão

		Predição	
		1	0
Real	1	Verdadeiro Positivo	Verdadeiro Negativo
	0	Falso Positivo	Falso Negativo

A tabela 1 nos mostra que se o eixo real e o eixo da predição forem 1, o resultado é um Verdadeiro Positivo; se o eixo real for 1 e o eixo da predição for 0, o resultado é um Verdadeiro Negativo; se e eixo real for 0 e o eixo da predição for 1, o resultado é um Falso Positivo; se eixo real e o eixo da predição forem 0, o resultado é um Falso Negativo(fn). Ela também nos fornece indicadores de desempenho como acurácia, precisão, *recall* e *f1-score* que podem ser calculados pelas formulas abaixo:

$$Acuracia = (VP + FN)/(VP + VN + FP + FN),$$

$$Preciso = VP/(VP + FP),$$

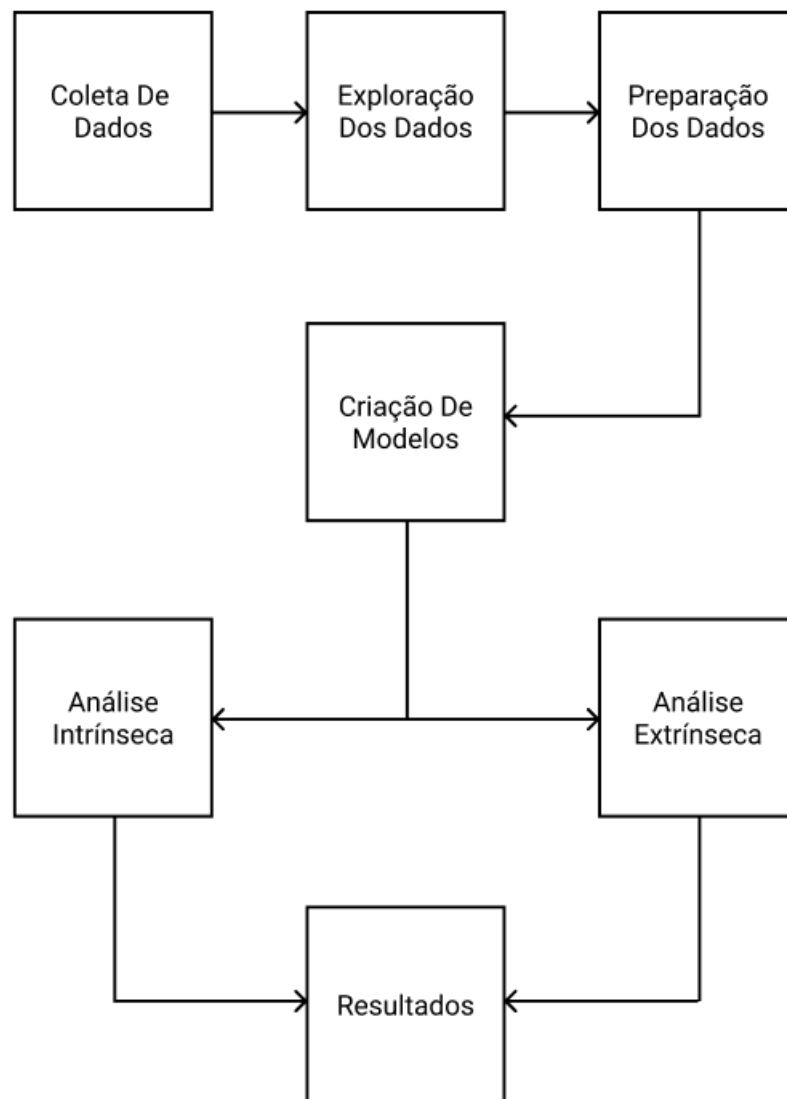
$$Recall = VP/(VP + FN),$$

$$f1-score = 2 * Preciso * Recall / (Preciso + Recall).$$

5 MATERIAIS E MÉTODOS

O método proposto para este trabalho pode ser identificado pelo diagrama de blocos da Figura 5. A partir de dados contendo, cotações diárias e de hora em hora, da Bitcoin, modelos serão treinados seguindo sua própria parametrização e posteriormente utilizados para fazerem previsões e serem analisados intrinsecamente e extrinsecamente. Os resultados serão então avaliados e comparados. A seguir serão descritos com detalhes cada um dos blocos e etapas do diagrama.

Figura 5 – Diagrama das etapas do projeto.



Fonte: Autoria Própria.

5.1 COLETA DE DADOS

A coleta dos dados foi feita através do site Kaggle (2020), para a cotação do Bitcoin no período de 08/10/2015 até 09/04/2020, tanto para o período diário (1D) quanto para o de hora em hora (1H). As variáveis encontradas nele correspondem a abertura, máxima, mínima e fechamento do preço no período determinado. Uma descrição sobre as variáveis pode ser visto a seguir na Tabela 2.

Tabela 2 – Descrição Das Variáveis

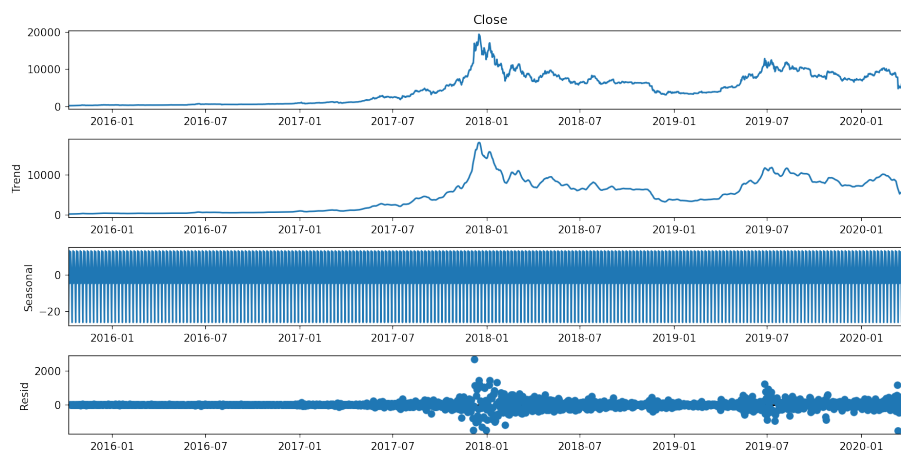
Variável	Descrição
Date	Dado da data e hora de gravação
Symbol	Simbolo do ativo de transação
Open	Valor em que as operações começaram
High	Maior valor atingido pelas operações
Low	Menor valor atingido pelas operações
Close	Valor em que as operações terminaram
Volume BTC	Volume transacionado em BTC
Volume USD	Volume transacionado em USD

Fonte: Autoria Própria

5.2 EXPLORAÇÃO DOS DADOS

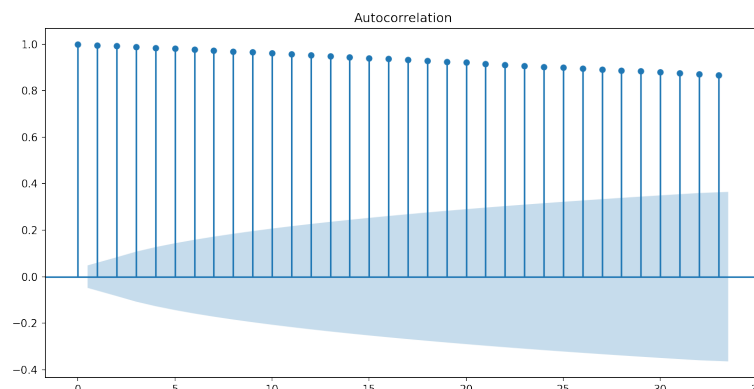
A exploração dos dados consistiu na decomposição e autocorrelação da série temporal que ajudou a extrair algumas informações úteis na construção dos modelos. As Figuras 6 e 7, mostram o que foi observado.

Figura 6 – Decomposição.



Fonte: Autoria própria.

Figura 7 – Autocorrelação.



Fonte: Autoria própria.

Na Figura 6 de decomposição foram observados 3 componentes: No componente de tendência (*Trend*) podemos observar 3 momentos principais, um de alta que vai desde 2016 até início de 2018, um de baixa que vai de 2018 até início de 2019, e um de alta novamente que vai de 2019 até o início de 2020, no componente de sazonalidade (*Seasonal*), observa-se que não há sazonalidade eminente e no componente de ruído (*Resid*) observou-se um ruído estável no início que vai aumentando até o fim da série, além de 2 aglomerações maiores em 2018-01 e 2019-07. Já na Figura 7 de autocorrelação, observa-se que há uma dependência até o *lag* 35, indicando que há necessidade de tornar a série estacionária.

5.3 PREPARAÇÃO DOS DADOS

Na preparação, os dados de fechamento (*Close*) foram separados em 80% para treino, de 08/10/2015 até 15/05/2019, e 20% para teste, de 16/05/2019 até 09/04/2020, avaliando assim os modelos com dados desconhecidos. Essa separação foi tanto para o diário, quanto para o de hora em hora. Somente o fechamento foi usado como entrada para os modelos, já que esse interessa mais aos investidores e negociadores. Como exemplo de como foi feita a separação temos a Figura 8:

Figura 8 – Separação dos dados. Em vermelho o treino e em azul o teste.

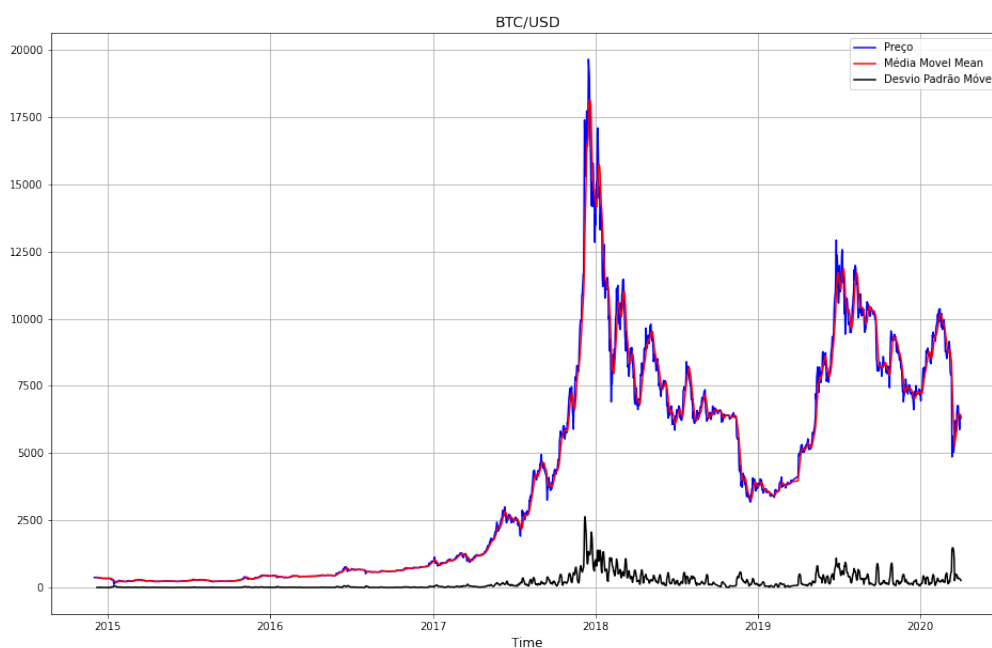


Fonte: Autoria própria.

onde a cor vermelha representa o treino, e a cor azul representa o teste.

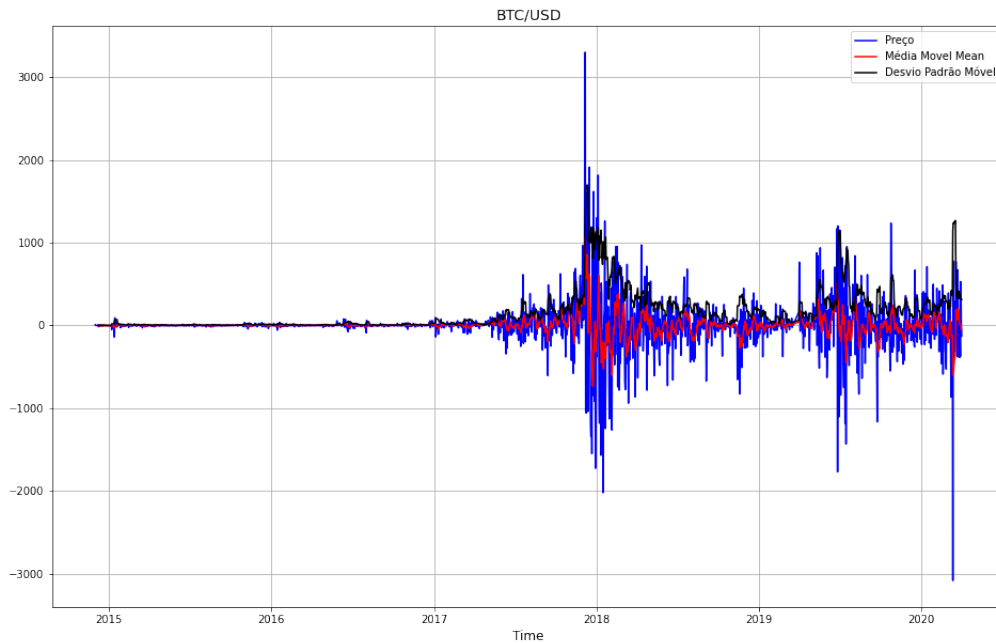
Já para lidar com a necessidade de tornar a série temporal estacionária, visto que os modelos ARIMA, MA, AR e Prophet possuem essa necessidade, foi utilizado o teste de *Augmented Dicky Fuller* (ADF).

Figura 9 – Média e desvio padrão do preço sem nenhuma diferenciação



Fonte: Autoria própria.

Figura 10 – Média e desvio padrão do preço com uma diferenciação



Fonte: Autoria própria.

O teste foi aplicado primeiro para os dados diários, sem nenhuma diferenciação com um $valor - P = 0.3832$, aceitando H_0 e concluindo assim que a série não é estacionária e portando precisa ser diferenciada. Na Figura 9, é possível visualizar como a média e o desvio padrão se comportaram ao longo da série para o primeiro teste. Para o segundo teste, foi feita 1 diferenciação e obteve-se um $valor - P = 0.0000001281$, recusando H_0 e concluindo assim que a série é estacionária e portando não precisa ser mais diferenciada. Na Figura 10, é possível visualizar como a média e o desvio padrão se comportaram ao longo da série, mostrando uma maior estabilidade, do que na figura do primeiro teste. Para os dados de hora em hora ocorreu de forma análoga, com a diferença de ' $valor - P = 0.59$ ' para o primeiro teste e ' $valor - P = 0$ ' para o segundo.

Por último, para facilitar o treinamento do LSTM, é importante que os dados estejam entre 0 e 1 e possuam a mesma escala. Para isso foi usado o método *MinMax* de normalização, dado pela equação a seguir:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (30)$$

onde z representa o valor normalizado, x representa o valor de entrada da série e, $\min(x)$ e $\max(x)$ representam os valores de mínimo e máximo da série.

5.4 CRIAÇÃO DE MODELOS

A criação de modelos corresponde ao treino e parametrização dos mesmos. Para os modelos AR, MA e ARIMA, seus parâmetros p , d , e q foram combinados variando de 0 a

5 e, através do estimador AIC, a melhor combinação foi escolhida. Também para o SES o estimador escolhido foi AIC e seu parâmetro α foi analisado variando de 0.05 até 0.95, com um passo de 0.05 por interação. Já o PROPHET foi utilizado de maneira a otimizar os parâmetros automaticamente pela sua biblioteca. O modelo *Naive* considera que a previsão ira repetir o último valor.

Para o modelo de aprendizado de máquina LSTM foram otimizados alguns parâmetros como: número de unidades de nós da camada e tamanho do lote de entrada, variando seus valores entre 5, 25, 50 e 100, e número de ciclos com seus valores variando de 50 até 400, com um passo de 50. O estimador para definir a melhor combinação dos valores foi o MSE. O modelo também usa o otimizador Adam, e a função de ativação Relu. Por fim, todos os modelos fizeram a previsão de 1 período a frente para todos os valores de testes.

5.5 ANÁLISE INTRÍNSECA

Após a previsão dos valores, foi realizada a análise intrínseca, que corresponde a métricas de regressão comparando os valores preditos pelos modelos e os valores reais. Nessa etapa foram aplicadas métricas MAPE, RMSE e R^2 para todas as previsões geradas pelos modelos, incluindo a *Naive*, para ambas as periodicidades.

5.6 ANÁLISE EXTRÍNSECA

A análise extrínseca, que corresponde na análise prática dos modelos criados por meio de sua incorporação em sistemas de compra e venda de Bitcoin, funcionou como da seguinte maneira:

- O primeiro sistema, chamado de Sistema Direcional, compara o preço da previsão com o preço atual, caso seja maior é feito a compra e caso, seja menor é feito a venda.
- O segundo sistema, chamado de Sistema de Votação, funciona como uma combinação do sistema anterior onde uma combinação de mais de 3 votos de todos os modelos são necessários para compra ou venda.
- O terceiro sistema, chamado de Sistema de *Buy and Hold*, foi utilizado como referência na comparação por se tratar de algo muito simples, comprando no início do período e vendendo no fim dele.

Todas as estratégias foram simuladas considerando um saldo inicial de U\$ 100 dólares e desconsiderando as taxas de operação.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesse capítulo são mostrados os resultados obtidos em nossos experimentos.

6.1 ANÁLISE INTRÍNSECA

Os resultados, referentes a análise intrínseca, considerando as métricas para cada modelo e sua periodicidade, podem ser vistos nas Tabelas 3, 4, 5 e 6, onde temos o modelo, MAPE, RMSE, R . Já para os melhores parâmetros encontrados em cada modelo para cada período temos as Tabela 7 e 8, onde a coluna "Melhores Parâmetros" representa um JSON com os melhores parâmetros encontrados.

Tabela 3 – Métricas Para Treino - Period. 1D

Modelo	MAPE	RMSE	R²
ARIMA	2.9123	282.8522	0.9942
AR	2.7595	287.3259	0.9940
MA	2.7509	287.8277	0.9940
PROPHET	inf	287.9446	0.0179
NAIVE	2.6204	290.6220	0.9939
SES	2.6192	293.2893	0.9937
LSTM	3.6419	319.4699	0.9926

Fonte: Autoria Própria

Tabela 4 – Métricas Para Previsão - Period. 1D

Modelo	MAPE	RMSE	R²
SES	2.9498	397.6040	0.9311
NAIVE	2.9779	400.8768	0.9299
PROPHET	3.0444	405.0639	0.9284
MA	3.0547	408.3656	0.9273
AR	3.0618	409.0631	0.9270
ARIMA	3.1188	411.5988	0.9261
LSTM	3.6327	445.1507	0.9136

Fonte: Autoria Própria

Tabela 5 – Métricas Para Treino - Period. 1H

Modelo	MAPE	RMSE	R²
ARIMA	0.4905	66.2537	0.9997
AR	0.4844	66.5128	0.9997
MA	0.4842	66.5248	0.9997
PROPHET	inf	66.6094	0.0013
NAIVE	0.4818	66.6538	0.9997
SES	0.4814	66.6845	0.9997
LSTM	0.7441	71.0544	0.9996

Fonte: Autoria Própria

Tabela 6 – Métricas Para Previsão - Period. 1H

Modelo	MAPE	RMSE	R²
SES	0.5044	82.8227	0.9970
MA	0.5046	82.8333	0.9970
AR	0.5047	82.8553	0.9970
ARIMA	0.5051	82.8410	0.9970
NAIVE	0.5058	83.0016	0.9970
PROPHET	0.5062	83.0408	0.9970
LSTM	0.6333	89.3332	0.9965

Fonte: Autoria Própria

Tabela 7 – Melhores Parâmetros Encontrados - Period. 1D

Modelo	Melhores parâmetros
AR	{"ordem_p": 5, "ordem_d": 1, "ordem_q": 0}
ARIMA	{"ordem_p": 4, "ordem_d": 1, "ordem_q": 5}
LSTM	{"lote": 5, "unidades": 100, "ciclos": 400}
MA	{"ordem_p": 0, "ordem_d": 1, "ordem_q": 5}
NAIVE	{}
PROPHET	{"ordem_d": 1}
SES	{"alfa": 0.9}

Fonte: Autoria Própria

Tabela 8 – Melhores Parâmetros Encontrados - Period. 1H

Modelo	Melhores parâmetros
AR	{"ordem_p": 4, "ordem_d": 1, "ordem_q": 0}
ARIMA	{"ordem_p": 4, "ordem_d": 1, "ordem_q": 3}
LSTM	{"lote": 100, "unidades": 100, "ciclos": 200}
MA	{"ordem_p": 0, "ordem_d": 1, "ordem_q": 2}
NAIVE	{}
PROPHET	{"ordem_d": 1}
SES	{"alfa": 0.9}

Fonte: Autoria Própria

Podemos notar nas tabelas 3 e 5, que as métricas de treino do modelo PROPHET, para ambas as periodicidades, apresentam um valor de "inf" para o MAPE, representando o valor de infinito, o qual acontece quando há divisões por zero ou próxima de zero e deve ser desconsiderado. Nota-se também que o modelo apresentou o menor valor de R^2 . Também para ambas as periodicidades, o modelo LSTM teve as piores métricas para treino de MAPE e RMSE.

Podemos notar na Tabela 4 que todas as métricas para previsão dos modelos com periodicidade diária se ajustaram pior que o modelo *Naive*, com os menores valores de MAPE e RMSE, e com os valores do R^2 mais próximos de 1, com exceção do modelo SES que conseguiu ter todas as métricas superiores não só a ele, mas a todos os outros. Já para a Tabela 6, temos quase que o inverso, com quase todos os modelos sendo superiores ao *Naive*, com exceção do Prophet e do LSTM, indicando assim que as previsões dos modelos se ajustaram muito melhor ao período de hora em hora do que o diário. Nota-se também que em ambos os casos o SES foi o que obteve as melhores métricas em comparação aos outros modelos.

Nas tabelas 7 e 8, devem ser feitas algumas considerações para os modelos *Naive* e Prophet, onde para o primeiro não há parâmetros para serem estimados por isso o JSON vem vazio e para o segundo, o parâmetro "ordem_d" na verdade, representa o tratamento da série temporal que precisou ser diferenciada para se ajustar melhor ao modelo, e não um parâmetro do modelo como acontece com o ARIMA, MA e AR.

Apesar das diferenças pequenas de desempenho dos modelos, todos atingem índices de correlação extremamente próximos de 100%, o que sugere grande precisão na predição, porém que, como apontado na seção a seguir, não se traduz necessariamente em competência prática.

6.2 ANÁLISE EXTRÍNSECA

Os resultados, referentes a análise extrínseca, considerando cada sistema utilizado podem ser vistos nas Tabelas 9, 11 e 13. Cada tabela nos mostra, como na Tabela 9 do Sistema Direcional, o desempenho de cada modelo para cada periodicidade, com seus ganhos, perdas,

quantidade de vezes que o modelo acertou e errou a direção prevista, porcentagem de acertos da direção prevista e balanço do saldo final (ganho - perda). Na Tabela 11 é ilustrado o desempenho do sistema de votação e na Tabela 13 temos o sistema de *Buy and hold*. Por último, nas Tabelas 10 e 12 temos detalhes da matriz de confusão e suas métricas para cada modelo e sua periodicidade.

Tabela 9 – Desempenho do Sistema Direcional para cada modelo testado

Modelo	Period.	Ganho	Perda	Acertos	Erros	% de acertos	Balanço
ARIMA	1D	425.26	-526.66	151	178	45.90	-101.40
MA	1D	437.21	-516.60	160	169	48.63	-79.39
PROPHET	1D	463.20	-512.38	151	178	45.90	-49.19
AR	1D	465.06	-492.26	155	174	47.11	-27.20
AR	1H	2014.70	-1971.12	4239	3654	53.71	43.58
LSTM	1D	511.95	-448.51	165	164	50.15	63.44
ARIMA	1H	2028.87	-1958.22	4208	3685	53.31	70.65
PROPHET	1H	2040.60	-1949.94	4013	3880	50.84	90.67
LSTM	1H	2067.04	-1924.82	4020	3873	50.93	142.22
SES	1D	570.55	-413.19	184	145	55.93	157.36
MA	1H	2096.14	-1894.62	4271	3622	54.11	201.51
SES	1H	2113.40	-1877.73	4327	3566	54.82	235.66

Fonte: Autoria Própria

Tabela 10 – Detalhes Do Sistema Direcional

Modelo	Period.	VP	FP	FN	VN	Acurácia	Precisão	Recall	F1-Score
LSTM	1D	1	163	1	164	0.5015	0.5000	0.0061	0.0120
LSTM	1H	27	3835	25	4006	0.5110	0.5192	0.0067	0.0132
AR	1D	73	82	82	92	0.5015	0.4710	0.4424	0.4562
MA	1D	76	84	80	89	0.5015	0.4872	0.4606	0.4735
PROPHET	1D	81	70	94	84	0.5015	0.4629	0.4909	0.4765
ARIMA	1D	83	77	87	82	0.5015	0.4882	0.5030	0.4955
ARIMA	1H	2152	2058	1820	1863	0.5087	0.5418	0.5360	0.5389
AR	1H	2154	2085	1788	1866	0.5093	0.5464	0.5358	0.5411
MA	1H	2153	2118	1759	1863	0.5088	0.5504	0.5361	0.5431
SES	1H	2165	2162	1714	1852	0.5089	0.5581	0.5390	0.5484
PROPHET	1H	2467	1552	2332	1542	0.5079	0.5141	0.6154	0.5602
SES	1D	94	90	74	71	0.5015	0.5595	0.5697	0.5646

Fonte: Autoria Própria

Tabela 11 – Desempenho do Sistema de Votação para cada modelo testado

Period.	Ganho	Perca	Acertos	Erros	% de acertos	Balanco
1D	388.06	-374.58	138	125	52.47	13.48
1H	1896.47	-1732.16	3784	3339	53.12	164.31

Fonte: Autoria Própria

Tabela 12 – Detalhes Do Sistema de Votação

Period.	VP	FP	FN	VN	Acurácia	Precisão	Recall	F1-Score
1D	47	91	45	80	0.4829	0.5109	0.3701	0.4292
1H	1708	2076	1542	1797	0.4921	0.5255	0.4873	0.5057

Fonte: Autoria Própria

Tabela 13 – Desempenho do Sistema de *Buy and hold* para cada modelo testado

Period.	Ganho	Perca	Acertos	Erros	% de acertos	Balanco
1D	494.29	-464.83	165	164	50.15	29.46
1H	2007.75	-1979.93	3997	3896	50.64	27.82

Fonte: Autoria Própria

Para o Sistema Direcional, como visto na Tabela 9, nota-se que o modelo SES tanto para a periodicidade diária quanto para a de hora em hora apresentou um balanço muito superior ao saldo inicial de U\$ 100 dólares. Nota-se também que apenas 2 modelos com periodicidade diária tiveram balanço positivo, LSTM e SES. Já para o Sistema de Votação, como visto na Tabela 11, observa-se que o balanço para a periodicidade diária teve um resultado pouco expressivo ao contrario do de hora em hora. Esse fenômeno já era esperado, uma vez que no Sistema Direcional a maioria dos modelos para essas periodicidades tiveram um balanço expressivo. Por último, no Sistema de *Buy and hold*, como visto na Tabela 13, vemos um resultado muito parecido para ambas as periodicidades.

Já para os detalhes do Sistema Direcional, como visto na Tabela 10, nota-se que todos tem uma acurácia muito próxima de 0.50, enquanto as outras métricas possuem valores mais variados. Nota-se também que o Recall e o F1-Score para o LSTM de ambas as periodicidades foram muito baixos em relação aos outros. Dentre os modelos testados, os que possuem os maiores valores para o F1-Score são o SES para ambas as periodicidades e o PROPHET(1H). Para os detalhes do Sistema de Votação, destaca-se também a periodicidade de hora em hora com valores superiores para todas as métricas.

Dentre os 3 sistemas de compra e venda, podemos observar que há 5 casos em que o balanço não conseguiu superar o Sistema de *Buy and hold*, sendo eles: 4 modelos do primeiro sistema ARIMA, MA, PROPHET e AR, e 1 caso do Sistema de Votação, ambos com periodicidade diária.

7 CONCLUSÃO

Neste trabalho foi apresentada uma comparação de alguns modelos estatísticos e de aprendizado de máquina e da criação de sistemas de compra e venda para a cotação do preço da Bitcoin. Para comparar os modelos e propor o sistema foi feita uma revisão sobre as criptomoedas, séries temporais, técnicas estatísticas e de aprendizado de máquina.

A metodologia proposta contou com o preço de fechamento do Bitcoin para ambas as periodicidades, e de técnicas estatísticas e de aprendizado de máquina como preditores. Foram apresentadas diferentes formas para avaliar os resultados de maneira intrínseca e extrínseca.

Os resultados mostram que os sistemas obtiveram um resultado melhor para a periodicidade de hora em hora, e um caso do Sistema Direcional se destaca entre eles, o modelo SES (1H) com \$235,66 de balanço final, com 54,82% de acerto e com 0.54 de *f1-score*, outros casos que também valem a pena se mencionar são: PROPHET (1H) com \$90,67 de balanço final, com 50,84% de acerto e com 0.56 de *f1-score*, SES (1D) com \$157,36 de balanço final, com 55,93% de acerto e com 0.56 de *f1-score* e MA(1H) com \$201,51 de balanço final, com 54,11% de acerto e com 0.54 de *f1-score*. Outro sistema de compra e venda, que se destacou foi o sistema de votação para a periodicidade de hora em hora, com \$164,31 de balanço final, com 53,12% de acerto e com 0.50 de *f1-score*. Entre os piores modelos destacam-se o ARIMA (1D) com \$-101,40 de balanço final, com 45,90% de acerto e com 0.49 de *f1-score*, MA (1D) com \$-79,39 de balanço final, com 48,63% de acerto e com 0.47 de *f1-score*, PROPHET(1D) com \$-49,19 de balanço final, com 45,90% de acerto e com 0.47 de *f1-score* e AR(1D) com \$-27,20 de balanço final, com 47,11% de acerto e com 0.45 de *f1-score*. Observa-se também um contraste entre análise intrínseca e extrínseca, em que apesar das métricas serem aproximadamente as mesmas para os modelos isso não se traduz necessariamente em boas políticas de compra e venda.

Sugestões para trabalhos futuros são:

- Utilização de outros bancos de dados, como por exemplo com periodicidade mensal, semanal e com dados mais recentes, para observar se os resultados continuam condizentes com os encontrados até agora e se há outras diferenças entre periodicidades;
- Extração de mais informações do que influencia a cotação do Bitcoin além de somente utilizar o preço de fechamento como entrada, aplicando maior representação de conhecimento e fornecendo um conjunto de medidas maior como entrada ao sistema;
- Desenvolvimento de uma interface para fácil utilização do sistema, ampliando seu uso customizado para a comunidade de investidores e estudantes.

Referências

- BIANCHI, F. M. et al. **Recurrent Neural Networks for Short-Term Load Forecasting An Overview and Comparative Analysis**. [S.l.]: Springer International Publishing, 2017. Citado na página 13.
- CHATFIELD, C. **The analysis of time series: an introduction**. 5. ed. Chapman and Hall/CRC, 1996. ISBN 0412716402,9780412716409. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=6544b01c91b53d8c283389551cca53d2>>. Citado na página 9.
- COINMARKETCAP. 2021. Disponível em: <<https://coinmarketcap.com/>>. Citado 2 vezes nas páginas 1 e 5.
- DENG, X. et al. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. **Information Sciences**, v. 340-341, p. 250–261, Jan 2016. Citado na página 16.
- HAMILTON, J. D. **Time Series Analysis**. 1. ed. Princeton University Press, 1994. ISBN 0691042896,9780691042893. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=452d247f10f10f3d402e98b16383461b>>. Citado 2 vezes nas páginas 9 e 11.
- HOUDT, G. V.; MOSQUERA, C.; NÁPOLES, G. A review on the long short-term memory model. **Artificial Intelligence Review**, v. 53, n. 8, p. 5929–5955, 2020. Citado na página 14.
- HYNDMAN ANNE B. KOEHLER, J. K. O. R. D. S. R. **Forecasting with exponential smoothing**. 1. ed. Springer, 2008. (Springer Series in Statistics). ISBN 3540719164,9783540719168,9783540719182. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=f27996ea9041a3bdad8606feec1114be>>. Citado na página 12.
- HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2nd. ed. Australia: OTexts, 2018. Citado na página 11.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679–688, 2006. Citado na página 15.
- INVESTING. 2020. Disponível em: <<https://br.investing.com/>>. Citado na página 8.
- KAGGLE. 2020. <https://www.kaggle.com/prasoonkottarathil/btcinUSD?select=BTCUSD_day.csv>. Accessed: 2020-06-30. Citado na página 18.
- KALYVAS, A. et al. What drives bitcoin's price crash risk? **Economics Letters**, Elsevier BV, v. 191, p. 108777, jun. 2020. Disponível em: <<https://doi.org/10.1016/j.econlet.2019.108777>>. Citado na página 1.
- KOTU, V.; DESHPANDE, B. Time series forecasting. In: **Data Science**. Elsevier, 2019. p. 395–445. Disponível em: <<https://doi.org/10.1016%2Fb978-0-12-814761-0.00012-5>>. Citado 3 vezes nas páginas 7, 8 e 9.
- LIANTONI, F.; AGUSTI, A. Forecasting bitcoin using double exponential smoothing method based on mean absolute percentage error. **JOIV : International Journal on Informatics Visualization**, v. 4, n. 2, p. 91, 2020. Citado na página 7.

- MCELREATH, R. **Statistical rethinking a Bayesian course with examples in R and Stan**. [S.l.]: CRC Press, Taylor amp; Francis Group, 2020. Citado na página 16.
- MCNALLY, S.; ROCHE, J.; CATON, S. Predicting the price of bitcoin using machine learning. **2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)**, 2018. Citado 2 vezes nas páginas 1 e 2.
- METATRADER4. 2021. Disponível em: <<https://www.metatrader4.com/>>. Citado na página 4.
- MIRJALILI, S.; FARIS, H.; ALJARAH, I. **Evolutionary machine learning techniques: algorithms and applications**. [S.l.]: Springer, 2020. Citado na página 13.
- MONTGOMERY CHERYL L. JENNINGS, M. K. D. C. **Introduction to Time Series Analysis and Forecasting (Wiley Series in Probability and Statistics)**. 1. ed. Wiley, 2008. ISBN 9780471653974. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=6bb1be93d7bca1e111d52d159e9246d2>>. Citado na página 11.
- NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. 2008. Citado 2 vezes nas páginas 1 e 6.
- NEVES, R. H. D. Bitcoin pricing: impact of attractiveness variables. **Financial Innovation**, v. 6, n. 1, 2020. Citado 2 vezes nas páginas 5 e 6.
- PAPADIMITRIOU, T.; GOGAS, P.; GKATZOGLOU, F. The evolution of the cryptocurrencies market: A complex networks approach. **Journal of Computational and Applied Mathematics**, v. 376, p. 112831, Oct 2020. Citado na página 5.
- PHALADISAILOED, T.; NUMNONDA, T. Machine learning models comparison for bitcoin price prediction. **2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)**, 2018. Citado na página 7.
- PROFILLIDIS, V.; BOTZORIS, G. Chapter 6 - trend projection and time series methods. In: PROFILLIDIS, V.; BOTZORIS, G. (Ed.). **Modeling of Transport Demand**. Elsevier, 2019. p. 225–270. ISBN 978-0-12-811513-8. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128115138000066>>. Citado na página 10.
- RIBEIRO, C. S. **Econometria**. 1. ed. Escolar Editora, 2014. ISBN 9789725924228. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=bb9af673d1de4f1c269daa6aa7ad89a6>>. Citado na página 10.
- ROY, S.; NANJIBA, S.; CHAKRABARTY, A. Bitcoin price forecasting using time series analysis. **2018 21st International Conference of Computer and Information Technology (ICCIT)**, Dec 2018. Citado 2 vezes nas páginas 1 e 7.
- SUBRAMANIAN, A.; SILVA, L. B. D.; COUTINHO, A. S. Aplicação de método e técnica multivariados para previsão de variáveis termoambientais e perceptivas. **Production**, v. 17, n. 1, p. 52–70, 2007. Citado na página 15.
- TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **Forecasting at Scale**, 2017. Citado 2 vezes nas páginas 12 e 13.
- WILKS, D. Time series. In: **International Geophysics**. Elsevier, 2011. p. 395–456. Disponível em: <<https://doi.org/10.1016%2Fb978-0-12-385022-5.00009-9>>. Citado na página 9.

YENIDOGAN, I. et al. Bitcoin forecasting using arima and prophet. **2018 3rd International Conference on Computer Science and Engineering (UBMK)**, 2018. Citado na página 7.

YU, Y. et al. A review of recurrent neural networks: Lstm cells and network architectures. **Neural Computation**, v. 31, n. 7, p. 1235–1270, 2019. Citado 2 vezes nas páginas 14 e 15.

ZULUTRADE. 2021. Disponível em: <<https://www.zulutrade.com/>>. Citado na página 4.