

Pré Processamento De Dados Para Análise de Séries Temporais na Previsão de Valores de Criptomoedas, Utilizando ARIMA e PROPHET

Adriano de Almeida Corrêa¹, Sylvio André Garcia Vieira¹

¹Sistemas de Informação – Universidade Franciscana
CEP 97.010-032 – Santa Maria – RS – Brazil

adrianoac12@gmail.com, sylvio@unifra.br

Abstract. *Cryptomania is a virtual currency, designed as a digital asset that uses cryptography to control the creation of units, check assets, and protect transactions. This study seeks to organize and analyze the daily closure value of the Bitcoin cryptomeo in the form of time series, in order to predict its future value. The methodologies of the ARIMA and PROPHET models were used to perform this prediction, being applied and demonstrated through the Python programming language. It is concluded that the ARIMA model constructed in this work generated a false prediction tendency, causing a misalignment with the real value of the currency, while the PROPHET model achieved better results, presenting a better prediction of values.*

Resumo. *Criptomoeda é uma moeda virtual, projetada como um ativo digital que usa criptografia para controlar a criação de unidades, verificar ativos, e proteger transações. Este estudo busca organizar e analisar o valor de fechamento diário da criptomoeda Bitcoin na forma de séries temporais, a fim de prever seu valor futuro. As metodologias dos modelos ARIMA e PROPHET foram utilizadas para realizar esta previsão, sendo aplicadas e demonstradas por meio da linguagem de programação Python. Conclui-se que, o modelo ARIMA construído neste trabalho gerou uma falsa tendência de previsão, ocasionando um desalinhamento com o valor real da moeda, enquanto o modelo PROPHET atingiu melhores resultados, apresentando uma melhor previsão de valores.*

1. Introdução

Um ponto fundamental para realizar uma análise técnica da Bolsa de Valores é entender sua repetição histórica. Baseado em seu histórico de cotações e negociações de ativos, é possível encontrar e identificar padrões [Jeske 2011]. Reconhecendo tais padrões de comportamento, e analisando seus dados na forma de séries temporais, pode ser possível prever valores futuros que auxiliam no processo de tomada de decisões para compra ou venda de ativos da Bolsa de Valores.

O Bitcoin foi a primeira moeda digital a operar por um meio descentralizado de troca digital, operando em um sistema distribuído ponto-a-ponto, sem um servidor centralizado ou ponto de controle [Nakamoto 2008]. De acordo com Bitcoin.org (2018), no ano de 2017, o enorme volume de transações a preços que não estavam de acordo com o valor intrínseco da moeda, resultou no que muitos chamaram de “Bolha Bitcoin”. Com a supervalorização no final de 2017 e desvalorização ocorrida no início de 2018, o valor

da moeda Bitcoin atingiu a estabilidade necessária para ser utilizada também como um ativo de investimento.

Desde então, a moeda Bitcoin sugere confiabilidade para ser anunciada no mercado de operações pela Intercontinental Exchange, operadora da bolsa de valores de Nova York (NYSE -The New York Stock Exchange) [Bitcoin 2018]. Um grupo de organizações, como BCG, Microsoft e Starbucks, pretende criar uma plataforma integrada para negociação e conversão de Bitcoin versus moedas fiduciárias¹, já que hoje o Bitcoin é a moeda digital com maior liquidez.

Os valores do Bitcoin podem ser considerados preços de ativos da Bolsa de Valores, e analisados em uma série temporal, para que por meio de métodos de modelagem como o ARIMA (Autoregressive Integrated Moving Averages), e PROPHET, seja possível prever seu comportamento futuro e assim identificar os melhores momentos para realizar operações de compra e venda.

Tendo em vista a utilização do mercado de ações como uma possibilidade de investimento a fim de obter um retorno financeiro positivo, deve-se considerar que este tipo de operação possui um alto risco de perda, desta forma, ter uma ferramenta que minimize riscos e indique uma estratégia de investimento com uma probabilidade maior de ganhos do que de perdas é muito válida.

1.1. Objetivo Geral

O presente trabalho tem por objetivo aplicar os modelos ARIMA e PROPHET sobre dados do mercado de criptomoedas, ajustando seus conjuntos de regras a fim de gerar previsões de preços futuros do Bitcoin.

1.2. Objetivos Específicos

O presente trabalho contempla os seguintes objetivos específicos:

- Estudar as características dos modelos ARIMA e PROPHET, para aplicar na série de valores dos preços diários de fechamento do Bitcoin;
- Organizar os dados históricos do Bitcoin, data e valor de fechamento, em séries temporais para serem analisados pelos modelos;
- Realizar testes com os modelos de previsão sobre estes dados organizados, selecionando o período de 1º de outubro de 2014 a 31 de agosto de 2018, para prever valores da primeira semana de setembro de 2018, dados estes já conhecidos, coletados em *CoinMarketCap*²;
- Avaliar os resultados obtidos e comparar com os dados já conhecidos, verificando a margem de erro dos resultados;
- Ajustar os cálculos dos modelos aplicados, para melhorar o desempenho na previsão de valores;
- Aplicar o modelo ajustado para realizar a previsão de valor da moeda;
- Comparar resultados de ambos os modelos, analisando a margem de erro de cada um, buscando identificar qual modelo obteve o melhor resultado.

¹ Moedas fiduciárias – Título não lastreado a nenhum metal (ouro, prata), não possuindo valor intrínseco (cheques, títulos de crédito, notas promissórias, entre outros) [Bancos 2018].

² CoinMarketCap – Mercado de Capitalização de Criptomoedas.

2. Referencial Teórico

Nesta seção serão abordados conceitos referentes ao modelo ARIMA e o modelo PROPHET, as tecnologias que contribuíram para o desenvolvimento deste trabalho, assim como os trabalhos correlatos que auxiliaram a pesquisa.

2.1. Séries Temporais

Uma série temporal é um conjunto de observações ordenadas em intervalo de tempo, ou seja, quando os dados são observados em diferentes instantes do tempo, podendo ser, diariamente como o preço de ações ou relatórios meteorológicos, mensalmente como a taxa de desemprego, ou trimestralmente como o PIB [Silva e Soeiro 2012].

Séries temporais possuem componentes como: Tendência, que pode aumentar ou diminuir a longo prazo (ex: apreciação do dólar vs real); Sazonalidade, flutuação periódica por determinado período na série temporal, formando um padrão que se repete (ex: aumento das vendas de cerveja no verão); Ciclicidade, comportamento que se repete após um grande intervalo de tempo, como meses, anos, etc.; Ruído ou Irregularidade, picos de valores em intervalos aleatórios em uma série temporal [Silva e Soeiro 2012]. A decomposição nos componentes de tendência, ciclicidade e sazonalidade, é a maneira tradicional de analisar uma série temporal.

Um fator que deve ser considerado em qualquer modelo estatístico de série temporal, é que a série a ser analisada deve ser estacionária, ou seja, ao longo de diferentes períodos de tempo deve manter média constante, variância constante ou desvio padrão, e a auto covariância não deve depender do tempo [Srivastava 2015]. Desta forma tendência e sazonalidade de uma série temporal precisam ser corrigidas por não apresentarem comportamento estacionário.

Uma série temporal precisa ser estacionária, pois sua constante facilita a aplicação em teorias e fórmulas matemáticas [Srivastava 2015]. Uma vez que a série temporal seja estacionária a precisão da previsão torna a probabilidade de que em intervalos diferentes ela mantenha o mesmo comportamento.

A estacionariedade da série temporal analisada pode ser verificada através de gráficos estatísticos (técnica visual), onde são apresentados valores de média móvel ou desvio padrão móvel para ver se variam ao longo do tempo. Existem também testes matemáticos que podem ser aplicados, auxiliando na identificação da estacionariedade, como o teste de Dickey-Fuller, que compara valores críticos para níveis de confiança [Srivastava 2015]. Se os dados do teste apresentarem valores menores que os valores críticos, a série é estacionária, do contrário é uma hipótese nula, ou seja, não é estacionária.

Ao analisar uma série temporal é possível compreender o mecanismo gerador da série, possibilitando obter e descrever a razão de seu comportamento, encontrar sua periodicidade, e controlar sua trajetória, além de predizer seu comportamento futuro, possibilitando fazer planos a longo, médio e curto prazo, e tomar decisões apropriadas [Migon 2007]. Estudando uma série temporal é possível:

- Analisar e modelar: descrever a série, verificar características relevantes, e possíveis relações com outras séries.

- Gerar previsão: utilizando valores da série temporal, e de outras séries se necessário e possível, são geradas previsões de curto prazo, onde os instantes previstos são chamados de horizonte de previsão.

2.1.1 Previsão de Séries Temporais

Para o estabelecimento de valores futuros de uma série temporal, é feita uma previsão com base tanto na informação atual, quanto a informações passadas. A margem de previsão está relacionada ao intervalo de tempo futuro, contado a partir de uma origem predeterminada [Esteves 2003].

Existem muitos procedimentos que podem ser utilizados no processo de previsão, podendo ser racionais e complexos, onde o trabalho de interpretação de uma série temporal exige um grande esforço; ou esse procedimento pode ser simples e intuitivo, necessitando de pouca análise dos dados.

A previsão é parte integrante de um processo de tomada de decisão, que possui alguns procedimentos estatísticos definidos pelos modelos [Gutiérrez 2003]:

- Modelos Univariados: modelos que se baseiam em uma única série histórica, como modelos de decomposição por componentes não observáveis, modelos de regressão, médias móveis, ajustamento sazonal e alisamento exponencial, e modelos de Box & Jenkins, que consistem em uma classe de modelos lineares chamados de modelos ARIMA;
- Modelos de Função de Transferência: a série analisada é explicada por seu histórico, e também por outras séries temporais não correlacionadas;
- Modelos Multivariados: utilizados para modelar duas ou mais séries, sem exigir causalidade em relação as séries temporais analisadas.

2.2. Modelos para Previsão de Séries Temporais

Nesta sessão são apresentados os conceitos dos modelos que serão utilizados para a previsão de valores baseados nas séries temporais.

2.2.1. ARIMA

O modelo ARIMA, cujo nome deriva do inglês AutoRegressive Integrated Moving Average, que significa auto regressivo integrado de média móvel, é um modelo muito utilizado para previsões e modelagem de séries temporais [Rocha 2015].

De acordo com este modelo, a série temporal é gerada por um processo estocástico³, cuja natureza pode ser reproduzida através de um modelo [Gutiérrez 2003], tendo como notação ARIMA (p,d,q), conhecido como abordagem de Box & Jenkins, a forma de representar a designação deste modelo, onde:

- p: representa o número de parâmetros auto-regressivos;
- d: número de diferenciações para a série tornar-se estacionária;
- q: número de parâmetros de médias móveis.

³ Processo estocástico – Família de variáveis aleatórias representando a evolução de um sistema de valores com o tempo.

Estes componentes representam respectivamente, AR(p), I(d), e MA(q), onde AR é a correlação entre os períodos de tempo anterior e atual, e a parte MA é usada para suavizar irregularidades [Srivastava 2015]. A parte I liga AR e MA, através do número de diferenciações da série.

Os termos auto-regressivos representam a série estacionária obtida por diferenciação, e as médias móveis as defasagens dos erros aleatórios, já o termo "integrado" refere-se ao processo de diferenciação da série original para torná-la estacionária [Gutiérrez 2003].

Através dos gráficos ACF (Função de Autocorrelação) e PACF (Função de Autocorrelação Parcial), é possível encontrar o valor de 'p' e 'q' para o ARIMA [Srivastava 2015].

- ACF em $y=0$, valor de 'q' determinado;
- PACF em $y=0$, valor de 'p' determinado.

O modelo ARIMA pela abordagem de Box & Jenkins, para fins de previsão de séries temporais, deve seguir as seguintes etapas [Gomes 1989], identificação, estimação, verificação e previsão.

Seguindo este ciclo definido, é necessário identificar o processo aleatório que gerou os dados, para assim estimar os parâmetros que o caracterizam, e verificar se foram cumpridas as hipóteses do modelo, se houver negativa na validação das hipóteses, uma nova fase de identificação deve ser realizada, até que a validação seja positiva, para então iniciar a realização da previsão da série temporal.

2.2.2. PROPHET

Prophet é um software de código aberto, desenvolvido para ser um modelo para a previsão de dados de uma série temporal, apresentando melhores resultados quando os dados analisados possuem uma periodicidade diária com pelo menos um ano de dados históricos. É empregado para produzir previsões para o planejamento e definição de metas, sendo muito confiável mesmo quando uma série temporal sofre uma variação brusca ao longo do tempo [Taylor e Letham 2017].

O modelo PROPHET oferece em sua formulação uma série de vantagens práticas, como [Taylor e Letham 2017]:

- Formulação flexível, é possível acomodar facilmente a sazonalidade com múltiplos períodos e diferentes pressupostos sobre as tendências;
- Ao contrário dos modelos ARIMA, as medidas da série temporal não precisam ter um período regular, sem necessidade de interpolar os valores faltantes;
- Permite que o analista explore de forma interativa muitas especificações do modelo;
- Possui parâmetros facilmente interpretáveis que podem ser alterados heurísticamente pelo analista para impor pressupostos na previsão.

Os analistas que fazem previsões geralmente possuem conhecimentos extensivos de domínio sobre o tema que estão prevendo, mas o conhecimento estatístico é limitado. A especificação do modelo de crescimento, de Taylor e Letham (2017), contempla vários

parâmetros onde o analista pode alterar o modelo para aplicar seus conhecimentos e conhecimento externo, sem exigir qualquer compreensão das estatísticas subjacentes.

2.3. Ferramentas Computacionais

Python é uma linguagem de programação de alto nível, de script, imperativa, orientada a objetos, interpretada e amplamente utilizada. A linguagem permite construir programas claros de pequena e grande escala. Possui um sistema de tipificação dinâmica, gerenciamento automático de memória e tem uma grande e abrangente biblioteca padrão [GSTI 2018].

Alguns aplicativos, permitem trabalhar com o Python, como Jupyter Notebook que é um aplicativo Web de código aberto que possibilita criar e compartilhar documentos que contêm código ativo, equações, visualizações e texto narrativo. Os usos incluem limpeza e transformação de dados, simulação numérica, modelagem estatística, visualização de dados, aprendizado de máquina [Jupyter 2018]. Neste trabalho será utilizado como ambiente para a linguagem Python.

Além disso, o Python trabalha com bibliotecas para auxiliar em determinadas tarefas, como o Pandas, que fornece estruturas de dados de alto desempenho, fáceis de usar e ferramentas de análise de dados [Pandas 2018].

Também foi utilizado neste trabalho o NumPy, pois é um pacote para computação científica para o Python. Sendo utilizado para trabalhar com objetos *array N-dimensional*, funções, álgebra linear, e números aleatórios. Além de seus usos científicos, o NumPy também pode ser usado como um contêiner multidimensional eficiente de dados genéricos [NumPy 2018].

O StatsModels é um módulo Python que fornece classes e funções para a estimativa de muitos modelos estatísticos diferentes, bem como para a realização de testes estatísticos e exploração de dados estatísticos [StatsModels 2018].

Para a representação gráfica foi utilizado o Matplotlib, uma biblioteca de plotagem 2D em Python que produz números de qualidade de publicação em uma variedade de formatos impressos e ambientes interativos entre plataformas [Matplotlib 2018].

3. Trabalhos Relacionados

Nesta seção serão apresentados estudos que possuem relação com a proposta, servindo de base para organização, motivação e definição do tema, a fim de buscar contribuições que auxiliaram no desenvolvimento deste trabalho.

3.1. Aplicação do Modelo ARIMA para Previsão do Preço do Frango Inteiro Resfriado no Grande Atacado do Estado de São Paulo

De acordo com Campos (2006), o estudo teve por objetivo elaborar um modelo de previsões para o preço do frango inteiro resfriado no grande estado de São Paulo, através da metodologia do modelo ARIMA. Utilizando uma série histórica mensal entre os anos de 1996 a 2005, realizou testes de previsão para os anos de 2004 e 2005. Após determinar 4 modelos que demonstraram consistência estatística para os testes, foi alcançado um bom desempenho de previsões.

3.2. Modelagem Box-Jenkins (ARIMA) para Gestão de Risco de Preço na Pecuária de Corte do Estado do Paraná

Em seu artigo, Padilha (2013) buscou ajustar um modelo de previsão ARIMA para a série temporal estacionária referente ao indicador de preços do boi gordo para o Estado do Paraná, visando auxiliar o setor da pecuária de corte a obter um referencial de preços futuros, permitindo assim uma eficiente gestão de risco de mercado.

3.2. A Guide to Time Series Forecasting with Prophet in Python 3

Em sua publicação, Vincent (2017) demonstra como visualizar e manipular dados de séries temporais para produzir previsões utilizando um novo método chamado PROPHET, desenvolvido pela equipe de *Core Data Science* do Facebook, permitindo que analistas de dados e desenvolvedores efetuem previsões em escala com a linguagem Python.

3.3. Considerações sobre os Trabalhos Relacionados

Os trabalhos correlatos foram pertinentes para a formulação da proposta inicial, possibilitando adquirir conhecimento para definir os modelos que serão utilizados para a predição de valores de criptomoedas.

No trabalho de Campos (2006), foram realizados testes de previsões trimestrais, anuais e para 2 anos, onde os modelos propostos apresentaram boa performance, surpreendendo com este resultado, visto que a metodologia ARIMA é reconhecidamente eficiente apenas para previsões de curto prazo.

O estudo de Padilha (2013) concluiu com base nos critérios estatísticos definidos, em uma previsão de sete dias, utilizando o modelo ARIMA, foram obtidos excelentes resultados, pois a diferença entre os valores reais observados e os valores estimados pelo modelo resultaram em uma margem de erro de 1,4%.

Pelo estudo de Vincent (2017), foi possível verificar como utilizar a biblioteca Prophet para executar a previsão de séries temporais em Python, onde foi demonstrado que é possível incluir e adaptar o modelo de previsão a diferentes requisitos para analisar uma série temporal.

4. Metodologia

Esta seção tem como objetivo descrever e aplicar os modelos ARIMA e PROPHET sobre a série temporal da criptomoeda Bitcoin, e a partir de suas técnicas ajustar os parâmetros dos modelos para a predição de valores. Os dados para análise foram coletados e organizados para serem analisados pela aplicação dos modelos ajustados.

Descrição das atividades:

- Levantamento de dados: coletar dados históricos da criptomoeda Bitcoin, definir quais variáveis serão observadas no estudo, e organizar estes dados para análise dos modelos;
- Estudo dos modelos: compreender como os modelos que serão aplicados devem ser parametrizados para analisar os dados da criptomoeda, a fim de gerar previsão de valores;

- Ajustar os modelos: aplicar os modelos para previsão sobre dados já conhecidos, verificando assim se os resultados obtidos são aceitáveis;
- Testes e ajustes finais: aplicar os modelos sobre diferentes intervalos de dados, com o intuito de identificar possíveis falhas e pontos críticos dos modelos;
- Prever valores futuros: gerar previsão de valores para intervalos futuros.

Visando obter uma previsão de curto prazo para os valores de fechamento diário da criptomoeda Bitcoin, foi necessário desenvolver uma pesquisa sobre a análise e previsão de séries temporais. Através desta análise buscou-se compreender como a modelagem do histórico de preços da criptomoeda poderia tornar-se um modelo de predição confiável.

Ao definir o ARIMA, e o PROPHET como meios para prever os valores futuros da criptomoeda, a linguagem Python foi escolhida para o desenvolvimento, análise e implementação dos modelos de previsão, por se tratar de uma ferramenta muito utilizada em análise de dados científicos, e também por possuir bibliotecas que implementam o ARIMA e o PROPHET.

O procedimento de previsão que envolve o valor de fechamento diário de uma criptomoeda, neste caso o Bitcoin, teve início pela seleção dos dados a serem analisados, seguido pelo desenvolvimento do modelo de previsão usando o método de média móvel integrada autorregressiva (ARIMA), e o modelo PROPHET.

Este estudo selecionou os dados diários de fechamento do Bitcoin a partir de 1º de outubro de 2014 até 31 de agosto de 2018. Estes dados são públicos e foram coletados em *CoinMarketCap* (2018).

As séries temporais tratam de duas colunas, uma é temporal, neste caso a data de fechamento, e outra é o valor a ser previsto, sendo aqui o valor da criptomoeda. Para facilitar a criação dos gráficos necessários para a análise da série temporal, a data de fechamento foi definida como índice destes dados. Na exibição e análise dos gráficos, o índice funcionará por padrão como o eixo x e, a coluna restante, será automaticamente considerada como o eixo y.

4.1. Aplicando o modelo ARIMA

Para a análise e predição de valores da série temporal da criptomoeda Bitcoin foram realizados os seguintes procedimentos:

- Visualização da série temporal, análise de dados que compõem a série;
- Tornar a série temporal em estacionária, caso não seja;
- Plotar os gráficos ACF/PACF para encontrar os parâmetros AR e MA;
- Construção do modelo ARIMA;
- Realização de previsão de valores da criptomoeda Bitcoin.

Ao visualizar os dados da série temporal, foi possível verificar que não se trata de uma série estacionária, além disso apresenta uma grande variação de valores, ascendente e descendente, entre o segundo semestre de 2017 e o primeiro semestre de 2018, conforme apresentado no Gráfico 1.

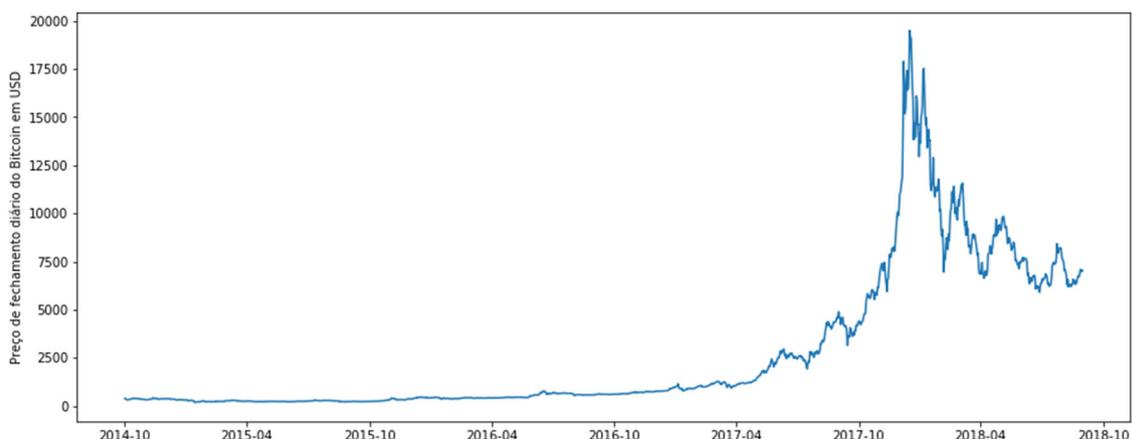
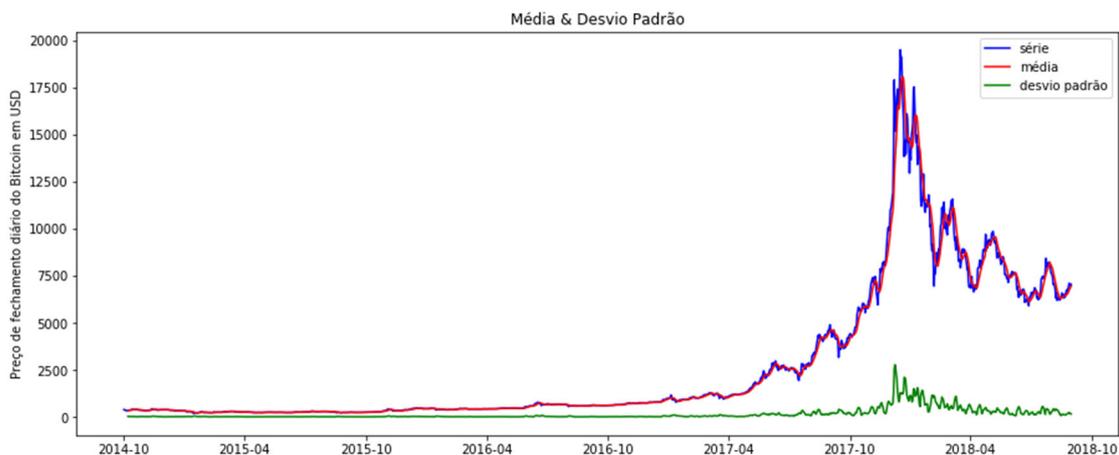


Gráfico 1. Preço de fechamento diário do Bitcoin em USD

Aplicando o teste Dickey-Fuller, os dados apresentados provam que a série não é estacionária, Gráfico 2, pois o valor do “Estatística de Teste” não é mais significativamente diferente de zero em relação ao “Valor Crítico” encontrado. É necessário então adequar a série temporal para torná-la estacionária. Os resultados do teste compreendem uma estatística de teste e alguns valores críticos para os níveis de confiança de diferença [Jain 2016]. Se a "estatística de teste" for menor que o "Valor crítico", pode-se rejeitar a hipótese nula e dizer que a série é estacionária.



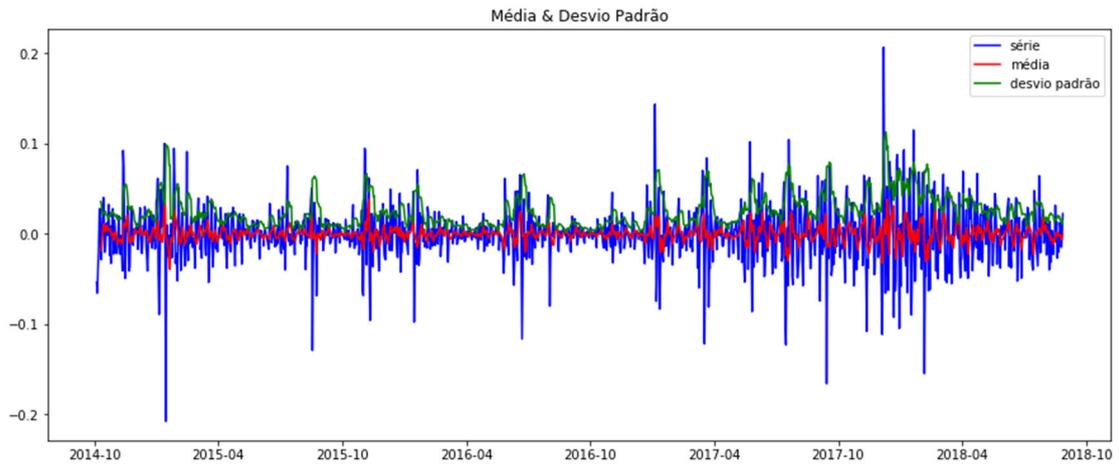
```

Resultados do Teste Dickey-Fuller:
Estatística de Teste      -1.697446
p-value                   0.432429
Atrasos                  22.000000
Número de Observações    1408.000000
Valor Crítico (1%)       -3.435003
Valor Crítico (5%)       -2.863595
Valor Crítico (10%)      -2.567864
dtype: float64

```

Gráfico 2. Resultados do Teste Dickey-Fuller

A fim de tornar a série estacionária foi aplicada uma transformação de log sobre os valores de fechamento do preço da criptomoeda, diminuindo a alteração exponencial da série, tornando a distribuição de valores de forma mais linear [Jain 2016]. Aplicando novamente o teste Dickey-Fuller sobre o conjunto de dados em log, são apresentados valores de média e desvio padrão mais similares, Gráfico 3.



Resultados do Teste Dickey-Fuller:

Estatística de Teste	-1.321111e+01
p-value	1.046585e-24
Atrasos	2.000000e+01
Número de Observações	1.404000e+03
Valor Crítico (1%)	-3.435016e+00
Valor Crítico (5%)	-2.863601e+00
Valor Crítico (10%)	-2.567867e+00
dtype:	float64

Gráfico 3. Série após transformação log

Após garantir a estacionariedade da série, foi possível determinar os valores de 'p' e 'q' do modelo ARIMA, através dos gráficos ACF e PACF.

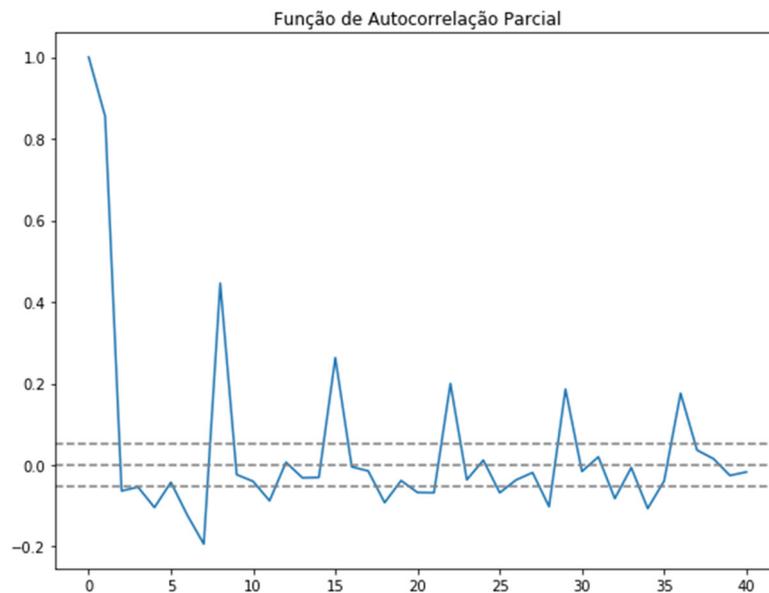


Gráfico 4. Gráfico da Função de Autocorrelação Parcial

No Gráfico 4 a faixa pontilhada em ambos os lados de 0 representam os intervalos de confiança da série, a análise deste gráfico determina o valor de 'p'.

- p - O valor de lag em que o gráfico PACF cruza o intervalo de confiança superior pela primeira vez. Neste caso $p = 2$.



Gráfico 5. Gráfico da Função de Autocorrelação

Observando o Gráfico 5, a faixa contínua em ambos os lados de 0 são os intervalos de confiança, sendo utilizado determinar o valor de ‘q’.

- q - O valor de lag no qual o gráfico ACF cruza o intervalo de confiança superior pela primeira vez. Neste caso, $q = 1$.

Definidos os parâmetros que serão utilizados na construção do modelo, neste caso definido como ARIMA(2, 1, 1), é possível aplicar sobre a série temporal e visualizar o resumo do mesmo. Através do resumo apresentado na Figura 1, é possível verificar dados como número de observações da série, possíveis variações para os valores de AR e MA, além de sua frequência.

```

=====
ARIMA Model Results
=====
Dep. Variable:          D.Close      No. Observations:      1430
Model:                 ARIMA(2, 1, 1)   Log Likelihood         2590.950
Method:                css-mle     S.D. of innovations     0.040
Date:                  Fri, 16 Nov 2018   AIC                    -5173.900
Time:                  16:08:18         BIC                    -5152.839
Sample:                10-02-2014       HQIC                   -5166.035
                    - 08-31-2018
=====
                    coef      std err      z      P>|z|      [0.025      0.975]
-----
ar.L1.D.Close      -0.4751      0.264      -1.800      0.072      -0.992      0.042
ar.L2.D.Close      -0.0245      0.028      -0.870      0.385      -0.080      0.031
ma.L1.D.Close       0.4929      0.263      1.875      0.061      -0.022      1.008
=====
Roots
=====
                    Real      Imaginary      Modulus      Frequency
-----
AR.1                -2.4031      +0.0000j      2.4031      0.5000
AR.2               -16.9513      +0.0000j     16.9513      0.5000
MA.1                -2.0289      +0.0000j      2.0289      0.5000
=====

```

Figura 1. Representação do Resumo do Modelo ARIMA

Apesar dos dados ‘p’ e ‘q’ serem definidos pela análise gráfica das funções ACF e PACF, podem ocorrer erros na representação do resumo do modelo ARIMA, pois os valores de lag selecionados podem causar uma interpretação de série não estacionária. Como solução simples o lag selecionado pode ser substituído por um lag próximo que esteja dentro dos parâmetros de construção do modelo.

Com o modelo construído foi possível efetuar previsões para os valores futuros da criptomoeda Bitcoin, porém neste estudo o intervalo a ser previsto já é conhecido, pois assim a comparação entre os valores previstos e os dados reais validam a confiabilidade da previsão. O intervalo escolhido para previsão foram os sete dias subsequentes em relação a série temporal estudada, ou seja, do dia 1º de setembro de 2018 a 7 de setembro de 2018.

4.2. Aplicando o modelo PROPHET

A implementação da análise da série temporal pelo modelo PROPHET é mais simplista e objetiva em comparação ao modelo ARIMA, pois a biblioteca do PROPHET espera a entrada de uma matriz contendo uma coluna com informações de tempo, e outra coluna contendo valores para previsão.

Como os dados já foram tratados na construção do modelo ARIMA, a série está adequada aos requisitos de importação pelo PROPHET, conforme demonstrado na Figura 2.

```
from fbprophet import Prophet # importação da biblioteca Prophet
```

```
dados_treino.head() # efetua leitura da série a ser analisada
```

```
Date
2014-10-01    383.61
2014-10-02    375.07
2014-10-03    359.51
2014-10-04    328.87
2014-10-05    320.51
Name: Close, dtype: float64
```

```
# cria uma nova base com os dados da base importada
dados_prophet = dados_treino.copy()
# informa ao Prophet a entrada das colunas tempo e métrica
dados_prophet = pd.DataFrame(dados_prophet)
# efetua indexação da série
dados_prophet.reset_index(drop=False, inplace=True)
# define o nome das colunas como 'ds' coluna tempo, e 'y' métrica a ser prevista
dados_prophet.columns = ['ds', 'y']
# exibe os dados da série
dados_prophet
```

	ds	y
0	2014-10-01	383.61
1	2014-10-02	375.07
2	2014-10-03	359.51
3	2014-10-04	328.87
4	2014-10-05	320.51
5	2014-10-06	330.08

Figura 2. Importação e adequação da série ao Modelo PROPHET

Após a importação da série pelo PROPHET, o mesmo impõe a nomenclatura para as colunas, sendo 'ds' para a coluna em função do tempo e 'y' para a coluna métrica. Neste ponto é possível exibir graficamente os valores e comportamento da série temporal, assim como exibir seus valores previstos.

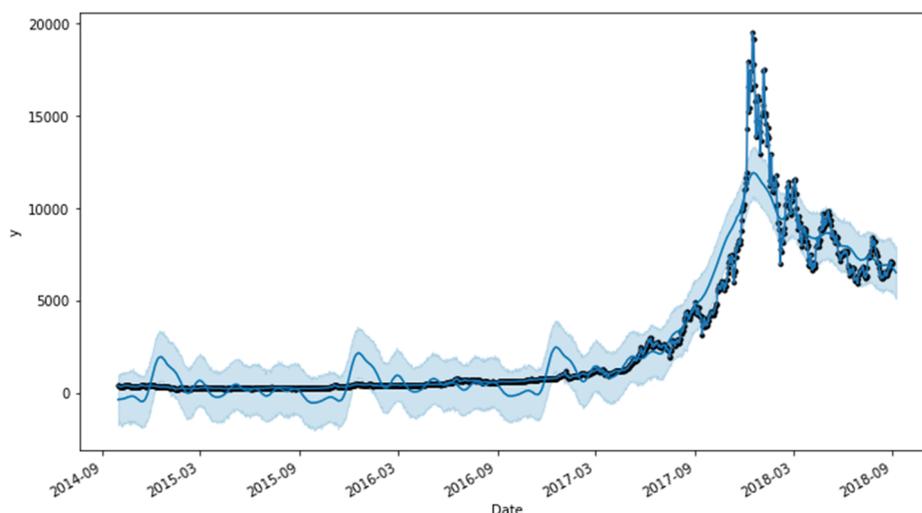


Gráfico 6. Demonstração gráfica da série pelo Modelo PROPHET

No Gráfico 6 é possível observar os valores analisados da série temporal (pontos pretos), a linha azul demonstra os valores previstos, e as regiões sombreadas representam intervalos de incerteza das previsões [Vincent 2017].

5. Resultados

Este trabalho apresentou uma proposta para previsão de valores da criptomoeda Bitcoin utilizando os modelos ARIMA e PROPHET, na tentativa de predição de valores futuros através da observação de seu histórico de comportamento.

Ao implementar o modelo ARIMA foi necessário analisar as funções de autocorrelação e autocorrelação parcial sobre o preço de fechamento do Bitcoin, porém a alta volatilidade da moeda determina a série como não estacionária. Para torná-la estacionária utilizou-se diferentes recursos como cálculo de diferenciação, transformação de logaritmo, extração de média ponderada exponencial, em um ciclo de teste de estacionariedade até que a série atingisse um padrão aceitável para realizar a análise pelo modelo.

Com os resultados obtidos por meio da análise da ACF e PACF foram determinados os valores para a construção do modelo ARIMA em (2,1,1), obtendo os valores futuros em relação ao preço de fechamento diário do Bitcoin, conforme demonstrado na Tabela 1.

Tabela 1. Margem de erro modelo ARIMA da criptomoeda Bitcoin

Data	Valor real	Valor previsto	% erro
01/09/2018	\$ 7.193,25	\$ 7.039,28	-2,14%
02/09/2018	\$ 7.272,72	\$ 7.037,01	-3,24%
03/09/2018	\$ 7.260,06	\$ 7.038,05	-3,06%
04/09/2018	\$ 7.361,66	\$ 7.037,61	-4,40%
05/09/2018	\$ 6.792,83	\$ 7.037,79	3,61%
06/09/2018	\$ 6.529,17	\$ 7.037,72	7,79%
07/09/2018	\$ 6.467,07	\$ 7.037,75	8,82%

Analisando a Tabela 1, onde os dados previstos foram gerados pelo modelo ARIMA, alcançou-se melhores resultados nos dois primeiros dias da previsão, onde os

valores da porcentagem de erro foram mais baixos. A partir do terceiro dia a variação dos valores previstos manteve-se estável não acompanhando a curva real dos valores da criptomoeda, observado no Gráfico 7.

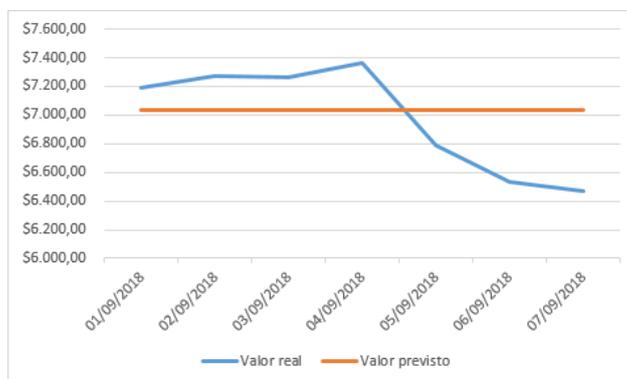


Gráfico 7. Valor real x Valor previsto, Modelo ARIMA

No modelo PROPHET a análise para previsão ocorre de maneira automatizada, em um loop de análise descrito por Taylor e Letham 2017, após a importação e ajuste da série pela biblioteca Prophet é possível efetuar a representação e predição dos valores da série imediatamente. A biblioteca Prophet efetua não somente a predição sobre o intervalo solicitado, mas de toda a série temporal, a fim de reconhecer qualquer tipo de padrão no comportamento dos dados analisados.

Considerando a frequência da série temporal, dados diários do fechamento do Bitcoin, define-se então que a variação do dia é a frequência desejada na análise e predição de valores pelo PROPHET, sendo assim foram previstos 7 dias futuros sobre a série para confrontar com os dados reais da moeda.

Tabela 2. Margem de erro modelo PROPHET da criptomoeda Bitcoin

Data	Valor real	Valor previsto	% erro
01/09/2018	\$ 7.193,25	\$ 6.763,36	-5,98%
02/09/2018	\$ 7.272,72	\$ 6.710,72	-7,73%
03/09/2018	\$ 7.260,06	\$ 6.678,90	-8,00%
04/09/2018	\$ 7.361,66	\$ 6.634,19	-9,88%
05/09/2018	\$ 6.792,83	\$ 6.583,09	-3,09%
06/09/2018	\$ 6.529,17	\$ 6.549,65	0,31%
07/09/2018	\$ 6.467,07	\$ 6.505,70	0,60%

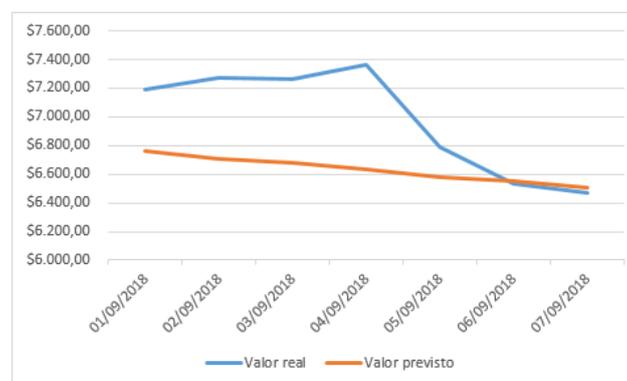


Gráfico 8. Valor real x Valor previsto, Modelo PROPHET

Observando o Gráfico 8, a tendência da previsão realizada pelo modelo PROPHET é confirmada a partir do sexto dia de previsão, onde as linhas “Valor real” e “Valor previsto” se encontram.

6. Conclusão e Trabalhos Futuros

Como conclusão, a abordagem de previsão utilizando o modelo ARIMA não produziu um modelo confiável de previsão, a variação dos valores previstos tornou-se estacionária e não conseguiu acompanhar o ambiente de alta volatilidade em que está inserida a criptomoeda Bitcoin.

Contudo o modelo PROPHET obteve sucesso ao indicar uma tendência de seus valores previstos, que se confirmou ao coincidir com a tendência dos valores reais do preço de fechamento do Bitcoin. Porém, para ambos os modelos se faz necessário considerações mais específicas quanto ao diagnóstico de erros.

Para uma prospecção futura de pesquisa, sugere-se uma modificação no modelo ARIMA, tornando possível analisar uma série temporal em conjunto de outra série temporal que possa ou não influenciar sua predição de valores, além de incluir dados de sazonalidade nas séries, transformando-se em modelo útil em casos onde os resíduos da série podem apresentar uma tendência ou padrão sazonal [Pereira 2016].

Ao PROPHET, como tema para futuros trabalhos, a adição de dados temporais sobre a série poderia trazer maiores benefícios no processo de predição do modelo, pois segundo Taylor e Letham (2017), o PROPHET é adequado para trabalhar com sazonalidades, mudanças históricas de tendências, feriados, ou curvas de crescimento não lineares.

Levando em consideração as diferentes implementações que podem ser feitas em ambos os modelos de predição utilizados, a pesquisa deste trabalho também pode ser ampliada para descobrir quais fatores impactam na volatilidade da taxa de câmbio de uma criptomoeda.

Referências

- Bitcoin, Portal do. (2018) “Maior Bolsa de Ações do Mundo Anuncia Mercado de Operações com Bitcoin”. Disponível em: <https://portaldobitcoin.com/maior-bolsa-de-acoes-do-mundo-anuncia-mercado-de-operacoes-com-bitcoin/>. Agosto.
- Campos, Paulo André Cavalcanti. (2006) “Aplicação do Modelo ARIMA para Previsão do Preço do Frango Inteiro Resfriado no Grande Atacado do Estado De São Paulo”. XIII Congresso Brasileiro de Custos – Belo Horizonte, MG, Brasil.
- CoinMarketCap. (2018) “Top 100 Cryptocurrencies by Market Capitalization”. Disponível em: <https://coinmarketcap.com/>. Setembro.
- Esteves, Gheisa Roberta Telles. (2003) “Modelos de Previsão de Carga de Curto Prazo”. Dissertação de Mestrado, Departamento de Engenharia Elétrica, PUC-Rio.
- Gomes, Francisco Carlos. (1989) “Os Modelos ARIMA e a Abordagem de Box-Jenkins, Uma Aplicação na Previsão do IBOVESPA a Curtíssimo Prazo”. Revista de Administração de Empresas. São Paulo, 29(2) 63-70. Abr./Jun. 1989.
- GSTI, Portal. (2018) “O que é Python?”. Disponível em: <https://www.portalgsti.com.br/python/sobre/>. Agosto.

- Gutiérrez, José Luis Carrasco. (2003) “Monitoramento da Instrumentação da Barragem de Corumbá-I por Redes Neurais e Modelos de Box & Jenkins”. Dissertação de Mestrado, Programa de Pós-graduação em Engenharia Civil da PUC-Rio.
- Jain, Aarshay. (2016) “A comprehensive beginner’s guide to create a Time Series Forecast”. Disponível em: <https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>. Agosto.
- Jeske, Jonas. (2011) “Similaridade de Séries Temporais na Bolsa de Valores”. Trabalho de Conclusão, Curso de Bacharelado em Ciência da Computação da UFRGS, Porto Alegre, BR-RS.
- Jupyter, Project. (2018) “The Jupyter Notebook”. Disponível em: <http://jupyter.org/>. Agosto.
- Matplotlib. (2018) “Matplotlib”. Disponível em: <https://matplotlib.org/>. Agosto.
- Migon, Hélio. (2007) "Análise de séries temporais." Departamento de Métodos Estatísticos, Instituto de Matemática, UFRJ, Disponível em: <http://acd.ufrj.br/~dani/pdf/slidespartefrequentista.pdf>. Junho.
- Nakamoto, Satoshi. (2008) “Bitcoin: A Peer-to-Peer Electronic Cash System”. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Julho.
- NumPy (2018) “NumPy”. Disponível em: <http://www.numpy.org/>. Agosto.
- Padilha Jr, João Batista. (2013) “Modelagem Box-Jenkins (ARIMA) para Gestão de Risco de Preço na Pecuária de Corte do Estado do Paraná”. X Convibra Administração – Congresso Virtual Brasileiro de Administração. Disponível em: <http://www.convibra.com.br/artigo.asp?ev=22&id=8283>.
- Pandas (2018) “Python Data Analysis Library”. Disponível em: <http://pandas.pydata.org/>. Agosto.
- Pereira, Cesar Machado. (2016) “Testes de Modelagens para Previsão de uma Série Temporal de Distribuição de Energia Elétrica”. XVIII Simpósio de Pesquisa Operacional & Logística da Marinha. Escola de Guerra Naval (EGN), Rio de Janeiro.
- Rocha, Vinícius Roggério da Rocha. (2015) “Modelo ARIMA”. Disponível em: <https://www.monolitonimbus.com.br/modelo-arima/>. Julho.
- Silva, Janaína Lima da e Soeiro, Joelma de Figueiredo. (2012) “Análise de tendência e sazonalidade da Série Temporal do Consumo de Energia Elétrica do Estado do Amapá”, Colegiado do Curso de Matemática, UNIFAP – Macapá – AP – Brasil.
- Srivastava, Tavish. (2015) “A Complete Tutorial on Time Series Modeling in R”. Disponível em: <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>. Junho.
- StatsModels (2018) “Welcome to Statsmodels’s Documentation”. Disponível em: <https://www.statsmodels.org/stable/index.html>. Agosto.
- Taylor, Sean J. e Letham Benjamin. (2017) “Forecasting at scale”. PeerJ Preprints 5:e3190v2. Disponível em: <https://doi.org/10.7287/peerj.preprints.3190v2>. Junho.
- Vincent, Thomas (2017) “A Guide to Time Series Forecasting with Prophet in Python 3”. Disponível em: <https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-prophet-in-python-3>. Junho.