

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

SVM aplicado a criptomoeda Ethereum

Claudio Vinicius Gonçalves

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Claudio Vinicius Gonçalves

SVM aplicado a criptomoeda Ethereum

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.
EXEMPLAR DE DEFESA

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Francisco Louzada Neto

USP – São Carlos
Agosto de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

G643s Gonçalves, Claudio Vinicius
 SVM aplicado a criptomoeda Ethereum / Claudio
 Vinicius Gonçalves; orientador Francisco Louzada
 Neto. -- São Carlos, 2019.
 62 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
 em Mestrado Profissional em Matemática, Estatística
 e Computação Aplicadas à Indústria) -- Instituto de
 Ciências Matemáticas e de Computação, Universidade
 de São Paulo, 2019.

 1. Aprendizado de Máquinas. 2. Maquinas de vetor
 de suporte. 3. Série Financeira. 4. Criptomoeda. I.
 Louzada Neto, Francisco, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Claudio Vinicius Gonçalves

SVM applied to criptocurrency Ethereum

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Francisco Louzada Neto

USP – São Carlos
August 2019

Dedico esse trabalho a minha mãe, Zelma. Sem ela, toda a inspiração, a ambição e a preparação para realização desse projeto não seriam possíveis.

AGRADECIMENTOS

Agradeço ao Prof. Dr. Anderson Ara, pela amizade e inspiração em iniciar esse projeto desafiador. Agradeço também ao Prof. Dr. Francisco Louzada Neto, que me suportou e aconselhou desde quando ainda era um aluno da graduação. Agradeço a minha mãe Zelma, que foi definitivamente o suporte e a força durante os momentos difíceis desse período. Agradeço aos meus amigos e familiares, que tiveram a paciência apoio, que veio de diversas maneiras, durante todo o curso. Agradeço a Ford, a empresa que confiou nessa empreitada desde o início desse desafio, ao Patrick, que desde o início me apoiou a continuar com esse projeto. E a todos os que cruzaram o meu caminho e me trouxeram inspiração e motivação, a vocês também deixou meu sincero agradecimento.

*“The key then to attaining this higher level of intelligence is to make our years of study
qualitatively rich.
We don’t simply absorb information - we internalize it and make it our own
by finding some way to put this knowledge to practical use.”
(Robert Greene)*

RESUMO

GONÇALVES, C. V. **SVM aplicado a criptomoeda Ethereum**. 2019. 62 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

A introdução da Blockchain e das criptomoedas trouxeram novas formas de realizar pagamentos, contratos, etc. Além desses benefícios, as criptomoedas também se tornaram opções de investimento para investidores. Esses investimentos apresentam grande risco, pois tem uma variabilidade muito alta, por exemplo a ETH, desde a sua criação, apresenta um crescimento de 13.000% desde 2014, onde foi oferta inicial da moeda. Esse projeto propõe a criação de modelos estatísticos clássicos e de aprendizado de máquina, para prever uma hora a frente o valor da criptomoeda. Para isso, inicialmente realizamos uma simulação com uma série de Lorenz, onde ajustamos um modelo ARMA e uma máquina de vetores de suporte com o Kernel Laplaciano para verificar a eficiência das técnicas. Iniciamos criando a série de Lorenz, ajustando os modelos e realizando previsões um passo a frente 24 vezes. Para verificarmos a qualidade do nosso modelo, comparamos esses valores preditos com os valores reais da série, e usamos como indicadores de qualidade as medidas RMSE e MAE, que na simulação resultaram para o modelo $ARMA(2, 3)$ $RMSE = 2,23$ e $MAE = 1,85$ e para o SVR $RMSE = 3,66$ e $MAE = 2,5$. Utilizando os valores preditos por ambos os modelos ajustados, buscamos melhorar a precisão combinando os modelos, obtendo assim para o SVR-ARMA $RMSE = 2,23$ e $MAE = 1,77$, melhorando assim a qualidade preditiva dos modelos. Seguimos aplicando para os dados reais as técnicas, inicialmente aplicando uma transformação log-retornos para a série de dados, e obtemos para o modelo $ARMA(2, 3)$ $RMSE = 2,23$ e $MAE = 1,85$ e para o SVR $RMSE = 3,15$ e $MAE = 2,4$, e quando combinamos os modelos temos o SVR-ARMA $RMSE = 2,08$ e $MAE = 1,75$.

Palavras-chave: SVR, Máquinas de Vetor de Suporte, Criptomoedas, Ethereum.

ABSTRACT

GONÇALVES, C. V. **SVM applied to cryptocurrency Ethereum**. 2019. 62 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

The introduction of Blockchain and cryptocurrency brought new ways to make payments, contracts, etc. Besides these benefits, the cryptocurrencies became an investment option for investors. These investments present great risk, due to high variability, for example ETH, since its creation, its growth reached 13.000% since 2014, when it was its Initial Coin Offer. This project proposes the creation of classical statistical models and machine learning, to predict one hour ahead the value of the cryptocurrency. For that purpose, initially we simulate a Lorez series, and fit a ARMA model and a SVM with a Laplacian Kernel to verify the technique efficiency. We create the simulated Lorez time series, adjust the model and calculate predictions one step ahead 24 times. For model quality we compare these predicted values with the real values of the series, and use as quality indicators RMSE and MAE, that in the simulation resulted for the $ARMA(2, 3)$ model a $RMSE = 2,23$ e $MAE = 1,85$ and the SVR $RMSE = 3,66$ and $MAE = 2,5$. Using both fitted models predicted values, to improve the prediction, we combine the models, resulting in an SVR-ARMA with $RMSE = 2,23$ and $MAE = 1,77$, improving the predictive quality of the models. We follow applying the techniques in the real data, firstly applying a log-returns transformation in the data, and adjusting the model $ARMA(2, 3)$ with $RMSE = 2,23$ and $MAE = 1,85$ and SVR with $RMSE = 3,15$ and $MAE = 2,4$, with the combine model SVR-ARMA with $RMSE = 2,08$ and $MAE = 1,75$.

Keywords: SVR, Support Vector Machine, Cryptocurrency, Ethereum.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação gráfica da função de perda suave para um SVM linear	38
Figura 2 – Cotação por hora	50
Figura 3 – Histograma do Fechamento	50
Figura 4 – Resultados da Simulação	52
Figura 5 – Atual vs Preditos (SVR, ARMA e SVR-ARMA)	53
Figura 6 – Série original da Ethereum e sua função de auto-correlação	54
Figura 7 – Série original da Ethereum e sua função de auto-correlação	54
Figura 8 – Série original da Ethereum e Valores preditos pelos modelos	55
Figura 9 – Amostra Teste e Valores preditos pelos modelos	55
Figura 10 – Amostra Teste e Valores preditos pelos modelos	56
Figura 11 – Ajuste ponderado SVM-ARMA	56

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas ETH por ano	48
Tabela 2 – Estatísticas descritivas ETH por mês	48
Tabela 3 – Estatísticas descritivas ETH por mês no ano 2018	49
Tabela 4 – Performance dos Modelos - Dados simulados	52
Tabela 5 – Performance dos Modelos - Dados reais	55

LISTA DE ABREVIATURAS E SIGLAS

AIC	Critério de informação de Akaike
AM	Aprendizado de máquina
ARIMA	Modelos Auto-Regressivos Integrados de Médias Móveis
BIC	Critério de Informação Bayesiano
ETH	Ether
FOREX	Foreign Exchange
MACD	Moving Average Convergence Divergence
MAE	Mean Absolute Error
MME	Média Móvel Exponencial
MMS	Médias Móveis Simples
RMSE	Root Mean Squared Error
RNA	Redes Neurais Artificiais
RSI	Índice de Força Relativa
SV	Vetores de Suporte
SVM	Máquinas de Vetores de Suporte
SVR	Vetores de Suporte para Regressão
TAE	Teoria de Aprendizado Estatístico

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Contextualizando	23
1.2	Definição do problema	25
1.3	Objetivos do Projeto	25
1.4	Organização dos capítulos	25
2	REVISÃO LITERÁRIA E TÉCNICAS CONVENCIONAIS	27
2.1	Trabalhos na área	27
2.2	Técnicas convencionais de Análise de Ativos	28
2.2.1	<i>Médias móveis</i>	28
2.2.2	<i>Candlesticks</i>	29
2.2.3	<i>Bandas de Bollinger</i>	29
2.2.4	<i>Índice de Força Relativa</i>	30
2.2.5	<i>Moving Average Convergence Divergence</i>	30
2.2.6	<i>Momentum</i>	31
2.3	Aprendizado de Máquina	31
2.3.1	<i>Máquinas de Vetores de Suporte</i>	31
2.4	Descrição das técnicas	32
3	MÉTODOS DE AVALIAÇÃO	33
3.1	Máquinas de Vetores de Suporte	33
3.1.0.1	<i>SVM de Margens Rígidas</i>	34
3.1.0.2	<i>SVM com Margens Suaves</i>	35
3.1.1	<i>SVMs não lineares</i>	36
3.1.2	<i>Máquinas de vetores de suporte para a regressão</i>	37
3.2	Modelos Estatísticos para Séries Temporais	40
3.2.1	<i>Modelos Lineares Estacionários</i>	41
3.2.2	<i>Processos auto-regressivos</i>	42
3.2.3	<i>Processo de Médias móveis</i>	43
3.2.4	<i>Modelos Auto Regressivos e de Médias móveis</i>	44
3.2.5	<i>Modelos ARIMA</i>	44
3.2.5.1	<i>Critério de informação</i>	45
3.3	Aplicação das técnicas	45

4	EXPERIMENTO E RESULTADOS	47
4.1	Análise Exploratória	47
4.2	Simulação de dados e comparação de modelos	51
4.3	Aplicação dos métodos para predição da Ethereum	53
4.4	Comentários sobre a Aplicação	57
5	CONSIDERAÇÕES FINAIS SOBRE O TRABALHO	59
	REFERÊNCIAS	61

INTRODUÇÃO

1.1 Contextualizando

Uma criptomoeda (ou cripto-moeda) é um ativo digital que funciona como um meio de troca que usa criptografia para assegurar suas transações, controlar a criação de novas unidades, e verificar a transferência dos ativos. Devido as suas características de segurança, criptomoedas não podem ser falsificadas. Uma das mais importantes características que definem as criptomoedas é a sua natureza orgânica, não sendo expedida por nenhuma autoridade central, deixando teoricamente imune a interferências governamentais ou manipulação. Assim, a criptomoeda é um tipo de moeda digital, se apresentando como uma alternativa das moedas e moedas virtuais. Assim, o controle descentralizado de cada criptomoeda funciona através da *Blockchain* que é uma base de dados de transações publicas, funcionando como um livro-razão distribuído (BHOSALE, 2018).

A primeira criptomoeda que teve maior aceitação publica foi a Bitcoin, que foi lançada em 2009 por um indivíduo ou grupo conhecido sob o pseudônimo de Satoshi Nakamoto. Esse individuo ou grupo publicou em 2008 o primeiro artigo apresentando sistema de pagamento eletrônico baseado em criptografia de prova, a *Blockchain*, ao invés de utilizar um terceiro de confiança, permitindo assim que duas partes interessadas possam transacionar diretamente entre si sem a necessidade desse terceiro de confiança (NAKAMOTO, 2008).

Em setembro de 2015, já existiam mais de 14,6 milhões de Bitcoins em circulação com o valor de mercado total de 3,4 bilhões de dólares. Devido ao sucesso da Bitcoin, outras alternativas competitivas de criptomoedas surgiram, como ETH e Litecoin (BHOSALE, 2018).

Temos hoje no mercado vários tipos de criptomoedas, sendo algumas delas:

- *Bitcoin* - A Bitcoin é uma criptomoeda e sistema de pagamento. A primeira criptomoeda descentralizada digital, que funciona sem nenhum banco ou administrador centralizador.

As transações são feitas pelos usuários diretamente, sem precisar de nenhum intermediário necessariamente. Essas transações são verificadas através de uma rede de nós que utiliza criptografia e são registradas em um livro-razão público chamado *Blockchain*. Ela foi inventada por um indivíduo ou grupo desconhecido que lançou a Bitcoin como um software de código aberto em 2009. Para incentivar a rede e a criação de nós, assim como assegurar a segurança da rede, as Bitcoins criadas são oferecidas como recompensa em um processo conhecido como mineração. As Bitcoins podem ser trocadas por outras moedas, produtos e serviços. Em Fevereiro de 2015, mais de 100.000 mercadores e vendedores aceitavam Bitcoin como pagamento. Uma pesquisa produzida pela Universidade de Cambridge estima que em 2017, existiam de 2.9 a 5.8 milhões de usuários únicos usando a carteira da criptomoeda, a maioria usando Bitcoin (BHOSALE, 2018).

- *Litecoin* - Litecoin é tratada como a maior rival da Bitcoin atualmente já que seu propósito foi desenhado para processar transações de menor valor de maneira rápida. Fundada em outubro de 2011 por Charles Lee, a diferença entre Bitcoin e Litecoin é que, para minerar Bitcoin o processamento é pesado exigindo mais da rede, já a Litecoin pode ser mineirada por um computador de mesa normal, com poder de processamento consideravelmente menor. Aproximadamente 84 milhões de litecoins estão em circulação em comparação com 21 milhões de Bitcoin, um outro ponto, o tempo de processamento de transação da Litecoin é de 2,5 minutos comparado com 10 minutos para a Bitcoin.
- *Etherium* - Conhecida também como Ether (ETH), essa criptomoeda é gerada pela plataforma Etherium, que é uma plataforma pública de código aberto baseado na blockchain. Ela trabalha em uma versão modificada da criada por Nakamoto, apresentando uma facilidade em contratos inteligentes. A Etherium foi inicialmente introduzida em 2013 por Vitalik Buterin, que foi um programador e pesquisador em criptomoedas. O desenvolvimento relacionado a Etherium foi fundado por uma venda pública online entre Julho e Agosto de 2014 e o desenvolvimento do sistema ficou online em 30 de julho de 2015. Teve inicialmente 11.9 milhões de moedas pré mineiradas para a venda pública. Essa circulação é quase 13% do total circulante da moeda (BHOSALE, 2018).

Considerando o alto risco relacionado a esse tipo de ativos, modelos preditivos podem ser ferramentas poderosas para prever o valor de determinada criptomoeda. Técnicas de análise convencionais nos mercados da Bolsa de Valores também podem ser aplicadas nesse mercado. Técnicas como *candlesticks*, médias móveis simples, médias móveis exponenciais, bandas de *Bollinger*, como outras, também podem ser usadas para análise visual e indicações de valorização ou desvalorização de determinada cripto-moeda.

1.2 Definição do problema

O investimento em tecnologias da blockchain incentiva uma nova era financeira, com serviços inovadores, seguros, rápidos, e sem interferência de terceiros. A criação de novas plataformas baseadas na blockchain acompanham o crescimento de novos criptoativos, que se tornam ativos atraentes para investidores ousados, devido a sua grande variabilidade e risco. Acompanhando o desenvolvimento científico em inteligência artificial e Aprendizado de máquina (AM), o desenvolvimento de ferramentas para o auxílio na previsão dos valores desses ativos com maior precisão e menor erro se torna uma oportunidade e necessidade nesse mercado de alto risco e variabilidade.

1.3 Objetivos do Projeto

Esse trabalho tem como objetivo aplicar técnicas de modelagem para comparar a performance de previsão de um modelo estatístico ARMA e técnicas de Vetores de Suporte para Regressão (SVR) com dados do valor de mercado na cotação base em dólar da moeda ETH.

1.4 Organização dos capítulos

A organização dessa dissertação vai seguir a seguinte ordem, no Capítulo 2 discutimos técnicas tradicionais de avaliação de ativos e trabalhos realizados nessa área, no Capítulo 3 desenvolvemos a metodologia escolhida. No Capítulo 4 descrevemos o conjunto de dados e aplicamos as metodologias estatísticas, discutimos os resultados obtidos da análise e, por fim, no Capítulo 5 concluímos a dissertação finalizando a discussão de resultados.

REVISÃO LITERÁRIA E TÉCNICAS CONVENCIONAIS

Neste capítulo vamos discutir alguns trabalhos realizados na modelagem e previsão de ativos financeiros, assim como descrever algumas técnicas tradicionais de análise de ativos.

2.1 Trabalhos na área

O artigo (NAKAMOTO, 2008) descreve a Bitcoin e introduz um sistema financeiro eletrônico *peer-to-peer* que permite transferências eletrônicas sem utilizar nenhum intermediário financeiro. Existem várias vantagens na utilização da Bitcoin, como velocidade de transação, segurança, custo e conveniência (RAYMAKERS, 2014). Blockchain é a tecnologia que suporta a Bitcoin, e só na primeira metade do ano de 2018 foi investido mais do que o total de 2017 (KPMG, 2018) na tecnologia. A Blockchain aumenta a eficiência e transparência de governança, financeira e acordos de segurança, e liberação de processos financeiros. Com origens em bases de dados distribuídas, os dados da blockchain são particionados em blocos, continuamente adicionando novos blocos sequenciais de dados (SWAN, 2015). Os blocos são ligados usando criptografias de assinaturas que resultam em transações com tempo registrado, a prova de adulteração. Estima-se que dentro de cinco anos a blockchain geraria uma economia de U\$16 bilhões por simplificar processos de auditoria e contabilidade. Trabalhando com uma versão modificada da tecnologia apresentada por Nakamoto, a Fundação Ethereum apresentou uma plataforma de contratos inteligentes ligada a criptomoeda ETH. O preço da Ethereum cresceu mais de 13.000% de 2014 a 2017 (BHOSALE, 2018).

A aplicação de aprendizado de máquina em Foreign Exchange (FOREX) vem ganhando popularidade nos últimos anos. (PENG, 2016) aplica modelos Máquinas de Vetores de Suporte (SVM) para 15 taxas de câmbio utilizando variáveis macroeconômicas usuais na literatura científica obtendo melhores resultados ao incorporar interações não-lineares com previsões mais

precisas e menor margem de erro.

Em um estudo coreano (CHEN KIHU JEONG, 2008) foi aplicado SVR comparado com outros três métodos, apresentando evidências que SVR para modelos ARMA apresenta uma excelente habilidade preditiva para dados reais e simulados de câmbio de dólar canadense contra dólar americano, demonstrando melhor performance que técnicas de Redes Neurais Artificiais (RNA) e modelos ARMA.

Segundo a dissertação de (PENG, 2016) a introdução de SVR teve êxito em fornecer maior precisão e menor margem de erro a previsões, devido a incorporação de interações não-lineares entre variáveis explicativas. Em seu estudo, (CHEN KIHU JEONG, 2008) apresentam evidências no uso redes *feedforward* e modelos SVM GARCH recorrente na previsão de erro de volatilidade e direção para um período a frente, performando bem também no caso de normalidade e amostras grandes. Ele acrescenta que a análise empírica é favorável a vantagem teórica do SVM. Em uma outra publicação (CHEN KIHU JEONG, 2010) mostra que a utilização de SVR melhora a performance de previsão de modelos *feedforward* padrão e tem melhor performance preditiva também do que em relação a estimação pela máxima-verossimilhança e RNA. (MAN-CHUNG WONG CHI-CHEONG, 2018) apresentam um modelo de redes neurais de 3 camadas para preços de ativos da bolsa comparando redes de queda ingrima e redes de gradiente conjugado, que mesmo apresentando o mesmo nível de erro e mesma precisão de previsão de direção, mostra que a eficiência da rede pode ser melhorada por conjugar um gradiente de aprendizado com inicialização e estimativa de pesos por uma regressão linear múltipla.

2.2 Técnicas convencionais de Análise de Ativos

Vamos descrever nessa sessão algumas das técnicas de análise convencional no mercado de trade.

2.2.1 Médias móveis

Uma Médias Móveis Simples (MMS) (MORETTIN, 2006) sobre n períodos é constituída das médias não ponderadas dos subconjuntos de N elementos p_1, \dots, p_n , onde $p(i)$ é o preço do ativo no i -ésimo período e n o número de períodos.

$$MMS = \frac{\sum_{i=1}^n p(i)}{n}. \quad (2.1)$$

Uma média móvel exponencial é muito utilizada como índice financeiro para medição de risco. As estimativas de risco são:

- Volatilidade - A partir do modelo de médias móveis, onde a série de retornos no tempo com n observações é ponderada por um fator de decaimento. A ponderação é feita de modo que observações mais recentes no tempo tem um peso maior que as observações mais remotas. O peso decai exponencialmente com n ;
- Correlação - São mais instáveis que as estimativas de volatilidade.

A Média Móvel Exponencial (MME) pode ser descrita da seguinte forma (MORETTIN, 2006):

$$MME_n(i) = \alpha * p(i) + (1 - \alpha) * EMA_n(i - 1), \quad (2.2)$$

$$\alpha = \frac{2}{1 + n}, \quad (2.3)$$

onde $p(i)$ é o preço do ativo no i -ésimo período e n o número de períodos.

2.2.2 *Candlesticks*

O gráfico de *candlesticks* representa o que ocorreu com o preço de uma ação dentro de um determinado período. Cada *candle* pode representar um minuto, cinco minutos, uma hora, uma semana, um mês ou até um ano (INVESTOPEDIA.COM, 2018). Cada *candle* é formado por valores importantes que o ativo alcançou dentro do período, sendo essas:

- Abertura - Valor de abertura do ativo;
- Fechamento - Valor de fechamento do ativo;
- Mínimo - O valor mínimo que o ativo teve durante o período;
- Máximo - O valor máximo que o ativo teve durante o período.

Estes valores compõem o *candle*, o corpo do *candle* trazendo informações de abertura e fechamento, e a sombra informando o mínimo e máximo. Outra característica da análise pelo gráfico de *candlesticks* é a sua cor, que determina se o *candlestick* é de alta ou de baixa em relação ao período anterior.

2.2.3 *Bandas de Bollinger*

As bandas *Bollinger* são constituídas por três curvas desenhadas em relação aos valores do ativo. A banda intermediária é uma medida de tendência intermediária, uma média móvel simples (MMS) geralmente, que serve como base para a banda superior e banda inferior.

O intervalo entre as bandas determinado volatilidade, geralmente representado pelo desvio padrão da mesma intervalo de tempo utilizado pela MMS, assim, quanto maior seu desvio padrão, maior a volatilidade de um ativo.

Assim, as bandas de *Bollinger* são descritas por (INVESTOPEDIA.COM, 2018):

- Banda intermediária = $MMS_{(n)}$;
- Banda superior = $MMS_{(n)} + DP_{(n)} * 0.9974$;
- Banda inferior = $MMS_{(n)} - DP_{(n)} * 0.9974$,

onde n é o número de períodos e DP é o desvio padrão.

2.2.4 Índice de Força Relativa

O Índice de Força Relativa (RSI) é um oscilador que mostra a força ou velocidade do preço de um ativo por meio de comparações individuais de movimentos crescentes ou decrescentes nos preços de fechamento. Esse indicador foi desenvolvido por Welles ([WILDER, 1978](#)). Para cada período, uma mudança crescente (U) ou decrescente (D) é calculada. "Período de queda" são caracterizados por um fechamento sendo menor do que o fechamento no período anterior e os períodos de alta por um fechamento maior do que no período anterior.

$$RSI = 100 - 100 \frac{1}{1 + RS}, \quad (2.4)$$

$$RS = \frac{MME_{(n)}U}{MME_{(n)}D}, \quad (2.5)$$

onde RS é a razão entre as média móveis exponenciais em n períodos da tendência de crescimento e decrescimento. Comumente se é usado um período de 14 intervalos para o RSI. O RSI está entre 0 e 100.

2.2.5 Moving Average Convergence Divergence

O Moving Average Convergence Divergence (MACD) é principalmente usado para identificar tendências de mudança. Ele é calculado a partir de MME de um período mais longo subtraído por um MME de um período mais curto. O MACD pode ser descrito como ([WANG, 2018](#)):

$$MACD_{(n)} = \sum_{i=1}^n EMA_k(i) - \sum_{i=1}^n EMA_d(i), \quad (2.6)$$

onde $d > k$.

2.2.6 Momentum

O momento é um indicador que mede a força de tendência de um ativo e é expresso em variação percentual do preço em um período de tempo estipulado. O *momentum* pode ser descrito pela diferença (ROSILLO J. GINER, 2018):

$$M = C - C_n, \quad (2.7)$$

onde M é o *Momentum*, C é o último preço cotado e C_n é o preço n cotado anterior tomado como referencia. Como padrão o período $n = 12$ é o mais utilizado. O *momentum* estuda a velocidade do movimento dos preços cotados em relação a um preço anterior e em muitos casos quando o preço ainda está em crescimento ou decrescimento, o *momentum* antecipa e muda, fazendo a próxima mudança uma tendência.

2.3 Aprendizado de Máquina

AM é uma disciplina focada em responder duas questões inter-relacionadas: Como podemos construir um constructo de sistemas computacionais que automaticamente melhore através da experiência, e quais são os fundamentos estatísticos-computacionais que governam todos os sistemas de aprendizado, incluindo computadores, humanos e organizações (MITCHELL, 2015). O estudo de aprendizado de máquina é importante tanto para endereçar essas perguntas, quanto pela alta praticidade computacional produzida nas suas áreas de aplicação. Assim, de acordo com Mitchell (MITCHELL, 1997) podemos então definir AM como "A capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência.". No AM, utilizando o princípio da inferência da indução, computadores são programados a aprender com experiências passadas. Sendo aplicadas em diversos problemas com sucesso, o AM vem se desenvolvendo trazendo novas técnicas e novos algoritmos cada vez mais eficientes e eficazes. Nesta seção será abordado de forma sucinta algumas técnicas.

2.3.1 Máquinas de Vetores de Suporte

As SVMs são baseados na teoria de aprendizado estatístico (VAPNIK, 1995). A Teoria de Aprendizado Estatístico (TAE)descreve os princípios a ser seguidos na obtenção de classificadores com boa capacidade de generalização em dado problema. A partir de um conjunto de dados de treinamento, seja \hat{h} um classificador, a TAE apresenta condições matemáticas que buscam \hat{h} levando em conta o desempenho do classificador no conjunto e sua complexidade. O objetivo é então obter um classificador com bom desempenho para novos dados dentro do mesmo domínio. Inicialmente assumimos que os dados domínio onde está ocorrendo o aprendizado são gerados de forma independente e identicamente distribuídas (i.i.d.). O erro esperado de um classificador h , também conhecido como risco esperado, para todos os dados desse domínio pode ser expresso pela equação 2.9,

$$R(h) = \int c(h(\mathbf{x}, y)) dP(\mathbf{x}, y), \quad (2.8)$$

$$c(h(\mathbf{x}, y)) = \frac{1}{2} |y - h(\mathbf{x})|. \quad (2.9)$$

$$(2.10)$$

Para minimizar o risco em 2.9, como a distribuição de probabilidade $P(\mathbf{x}, y)$ é desconhecida, utiliza-se o princípio da indução com o objetivo de inferir uma função \hat{h} que minimize o erro sobre os dados de treinamento e também leve o menor erro para novos dados. Para medir o desempenho de um classificador, temos que o risco empírico de h é dado por meio da taxa de classificações incorretas,

$$R_{emp}(h) = \frac{1}{n} \sum_{t=1}^n c(h(\mathbf{x}_i, y_i)). \quad (2.11)$$

Com base nos dados de treinamento conhecidos, esse processo de indução constitui o princípio de minimização do risco empírico (SMOLA, 2002). Assintoticamente, com $n \rightarrow \infty$, podemos estabelecer condições para o algoritmo de aprendizado que garantam a obtenção de classificadores com o risco empírico convergindo para o risco esperado (MULLER K. R., 2001). Para evitar *overfitting*, e que se tornem pouco informativos, é preciso restringir a classe de funções da qual \hat{h} é extraída. Nesse trabalho utilizaremos metodologias para SVR.

2.4 Descrição das técnicas

Nesse capítulo discutimos algumas técnicas tradicionais utilizadas no mercado financeiro de investimentos de ativos, assim como discutimos alguns trabalhos importantes realizados nessa área com o objetivo de prever valores da bolsa de valores e mercados similares. Assim, nesse trabalho vamos estudar a eficiência na previsão de valores futuros da cotação do cripto-ativo ETH utilizando uma abordagem de aprendizado de máquina e um modelo clássico estatístico. Para esse objetivo, no próximo capítulo vamos descrever as técnicas que serão aplicadas nesse trabalho.

MÉTODOS DE AVALIAÇÃO

Nesse capítulo vamos descrever com mais detalhes as técnicas que serão utilizadas nesse trabalho, sendo elas a Máquina de Vetor de Suporte para Regressão e Modelos clássicos de série temporal ARIMA.

3.1 Máquinas de Vetores de Suporte

Introduzido pela teoria de aprendizado estatístico de Vapnik (VAPNIK, 1995), (VAPNIK, 1997), o SVM são um exemplo de classificadores lineares de duas classes. Sejam h um classificador e H o conjunto de todos os classificadores de um processo de aprendizado de máquina. Dado um conjunto de treinamento X , composto de n pares (\mathbf{x}_i, y_i) , para gerar um classificador particular $\hat{h} \in H$. Para escolha do classificador devemos supor que os dados do domínio são gerados de forma independente e identicamente distribuídas (i.i.d) de acordo com a distribuição de probabilidade $P(\mathbf{x}, y)$ que descreve a relação entre os objetos e seus rótulos. O erro (risco) esperado de um classificador h para esse domínio é dado por:

$$R(h) = \int c(h(\mathbf{x}, y)) dP(\mathbf{x}, y), \quad (3.1)$$

que mede a capacidade da generalização de h . Essa é função de custo relacionada a previsão $h(\mathbf{x})$ para a saída y , onde $y_i \in \{-1, 1\}$. Um tipo de função usada em problemas de classificação é a função 0-1 definida,

$$c(h(\mathbf{x}, y)) = \frac{1}{2} |y - h(\mathbf{x})|, \quad (3.2)$$

onde essa função retorna $\mathbf{x} = 0$ se classificado corretamente e $\mathbf{x} = 1$ caso contrário. Como não é possível minimizar o risco esperado diretamente, já que a distribuição de probabilidade dos dados é desconhecida. Utilizando a informação dos dados de treinamento, através do princípio da indução é possível inferir uma função \hat{h} que minimize o erro sobre os dados de treinamento

e assim, esperar que também leve o menor erro no procedimento para os novos dados. O risco empírico de h é dado por,

$$R_{emp}(h) = \frac{1}{n} \sum_{i=1}^n c(h(x_i), y_i). \quad (3.3)$$

Esse processo de indução constitui o princípio de minimização do risco empírico (SMOLA, 1998), que com $n \rightarrow \infty$, assintoticamente, é possível estabelecer condições para que o risco empírico converta para o risco esperado.

3.1.0.1 SVM de Margens Rígidas

Dado um conjunto de dados linearmente separável, SVMs lineares com margens rígidas definem fronteiras lineares para esses dados. Classificadores lineares separam dados por meio de um hiperplano, que pode ser descrito por (SMOLA, 1998),

$$h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b, \quad (3.4)$$

onde $\mathbf{w} \cdot \mathbf{x}$ é o produto escalar entre os vetores \mathbf{w} e \mathbf{x} , $w \in X$. Essa equação pode ser usada para dividir o espaço X em duas regiões, $\mathbf{w} \cdot \mathbf{x} + b > 0$ e $\mathbf{w} \cdot \mathbf{x} + b < 0$. A partir de $h(\mathbf{x})$ é possível obter um número infinito de hiperplanos equivalentes, a partir de uma multiplicação de \mathbf{w} e b por uma constante. Define-se como hiperplano canônico em relação ao conjunto \mathbf{X} como aquele em que \mathbf{w} e b mais próximo do hiperplano $\mathbf{w} \cdot \mathbf{x} + b = 0$ satisfaçam:

$$|\mathbf{w} \cdot \mathbf{x}_i + b| = 1, \quad (3.5)$$

implicando nas inequações,

$$\begin{cases} \mathbf{w} \cdot \mathbf{x} + b \geq +1 & \text{se } y_i = +1, \\ \mathbf{w} \cdot \mathbf{x} + b \leq -1 & \text{se } y_i = -1, \end{cases}$$

$$y_i(\mathbf{w} \cdot \mathbf{x} + b) - 1 \geq 0, \forall (\mathbf{x}_i, y_i) \in \mathbf{X}. \quad (3.6)$$

Para resolver esse problema devemos minimizar $\|\mathbf{w}\|$, que pode ser resolvido pelo seguinte problema de otimização.

$$\text{Minimizar } \frac{1}{2} \|\mathbf{w}\|^2,$$

restrito a $y_i(\mathbf{w} \cdot \mathbf{x} + b) - 1 \geq 0$. Esse problema é quadrático e pode ser resolvido pela função lagrangiana que engloba as restrições da função objetivo (SMOLA, 2002),

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i(\mathbf{w} \cdot \mathbf{x} + b) - 1), \quad (3.7)$$

e obtemos o seguinte problema de otimização,

$$\text{Maximizar}_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j),$$

$$\text{Com restrições} \begin{cases} \alpha_i \geq 0, & \forall i = 1, \dots, n, \\ \sum \alpha_i y_i = 0, \end{cases}$$

que é denominado de forma dual, que apresenta restrições mais simples e permite representar o problema de otimização em termos de produtos internos entre os objetos. O ponto ótimo para o produto entre as variáveis duais e as restrições deve ser nulo (SMOLA, 1998), assim temos:

$$\alpha_i^* (y_i (\mathbf{w}^* \cdot \mathbf{x}_i + b^*) - 1) = 0, \forall i = 1, \dots, n. \quad (3.8)$$

Pela equação, α_i^* é diferente de zero somente para objetos que se encontram sobre os hiperplanos H_1 e H_2 . Onde $\alpha_i^* > 0$ temos os vetores de suporte Vetores de Suporte (SV) e são objetos mais informativos do conjunto de treinamento (BURGES, 1998). O valor de b^* é calculado a partir das SV's,

$$b^* = \frac{1}{n_{SV}} \sum_{x_j \in SV} \frac{1}{y_i} - \mathbf{w}^* \cdot \mathbf{x}_j = \frac{1}{n_{SV}} \sum_{x_j \in SV} \left(\frac{1}{y_i} \alpha_i^* y_i \mathbf{x}_i \cdot \mathbf{x}_j \right), \quad (3.9)$$

onde n_{SV} é o número de SV's e SV representa o conjunto de SVs. E a partir disso, tem-se o classificador $g(\mathbf{x})$ dado por,

$$g(\mathbf{x}) = \text{sgn}(h(g(\mathbf{x}))) = \text{sgn} \left(\sum_{x_j \in SV} y_i \alpha_i^* \mathbf{x}_i \cdot \mathbf{x}_j + b^* \right). \quad (3.10)$$

3.1.0.2 SVM com Margens Suaves

Em situações onde os dados não podem ser linearmente separáveis, seja por outliers ou ruídos, ou pela natureza do problema que pode não ser linear. Para lidar com esse problema utilizamos variáveis de folga ξ_i para todo $i = 1, \dots, n$. Essas variáveis servem para relaxar as restrições impostas ao problema de otimização primal (CHEN KIHONG, 2010),

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, n.$$

Aplicando esse procedimento podemos suavizar as margens do classificador linear e permitir a ocorrência de alguns erros de classificação. Para minimizar os erros sob os dados de treinamento, temos,

$$\text{Minimizar}_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum_{i=1}^n \xi_i \right),$$

onde C é uma constante que impõe um peso ao minimizar os erros do conjunto de dados de treinamento. Assim temos o problema dual,

$$\text{Maximizar}_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j),$$

$$\text{Com restrições} \begin{cases} 0 \geq \alpha_i \geq C, & \forall i = 1, \dots, n, \\ \sum_{i=1}^n \alpha_i y_i = 0, \end{cases}$$

que com relação ao problema em margens rígidas, só temos diferença na restrição em α_i que agora são limitados pelo valor de C . As variáveis ξ_i^* podem ser calculadas por (SMOLA, 1998),

$$\xi_i^* = \max\{0, 1 - y_i \sum_{j=1}^n y_j \alpha_j^* x_j \cdot x_i + b^*\}, \quad (3.11)$$

e b^* é calculado a partir de α^* ,

$$\alpha_i^* (y_i (\mathbf{w}^* \cdot \mathbf{x}_i + b^*) - 1 + \xi_i^*) = 0, \quad (3.12)$$

$$(C - \alpha_i^*) \xi_i^* = 0. \quad (3.13)$$

3.1.1 SVMs não lineares

SVMs não lineares mapeiam o conjunto de treinamento no seu espaço original (entradas), para um espaço de maior dimensão (espaço característica) (FACELI ANA CAROLINA LORENA, 2011). A ideia é, seja $\Phi: X \rightarrow \chi$ um mapeamento, em que X representa o espaço de entrada e χ o espaço de características, devemos escolher um Φ que faz com que o conjuntos de treinamento mapeado em χ possa ser separado por uma SVM linear. Motivado pelo teorema de Cover (HAYKIN, 1999), dado o conjunto não linear de espaço de entrada X , podemos transformar em um espaço características com alta probabilidade dos objetos serem separados linearmente. Como o espaço característica pode ter dimensão muito grande, ou até mesmo infinita, a computação de Φ pode ser muito custosa ou inviável. Para isso utilizamos funções kernel, que realizam o produto escalar entre os objetos do espaço característica. Um kernel K é uma função que recebe dois pontos x_i e x_j no espaço de entradas e calcula o produto escalar desses objetos no espaço de características (FACELI ANA CAROLINA LORENA, 2011),

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j). \quad (3.14)$$

A utilidade dos kernels está na sua simplicidade de calcular e sua capacidade de representar espaços abstratos.

3.1.2 Máquinas de vetores de suporte para a regressão

A ideia da Máquina de vetores de suporte para a regressão é, dado $\{(x_1, y_1), \dots, (x_{tr}, y_{tr})\} \subset \mathcal{X} \times \mathbb{R}$ um conjunto de dados de treinamento dentro do espaço entrada \mathcal{X} , como valores de cotação de certo ativo, por um período t de unidade de tempo (minutos, horas, dias, etc) e outros indicadores econômicos. O objetivo então é encontrar uma função $h(x)$ que produza saídas contínuas para os dados de treinamento que desviem no máximo de ε de seu rótulo desejado sendo o mais uniforme e regular possível (VAPNIK, 1995). Assim, como o nome sugere, o SVR se baseia em extrair um subconjunto dos dados de treinamento que servem como vetores de suporte que buscam representar uma característica estável dos dados. Em problemas de classificação, a variável y é binária, -1 e 1 , enquanto na regressão y pode tomar qualquer valor real. Buscamos achar função $f(x)$ que tenha um desvio ε no máximo de y , para todo o conjunto de dados de treinamento. Na forma linear, $f(x)$ é dado por (SMOLA, 2002),

$$f(x) = w \cdot x + b, \quad (3.15)$$

onde $w \in \mathcal{X}, b \in \mathbb{R}$ (SMOLA, 2002) e $\langle w \cdot x \rangle$ é o produto interno em \mathcal{X} . Onde podemos encontrar $f(x)$ minimizando w através da norma (SMOLA, 1998) $\|w\|^2 = w \cdot w$. Isso pode ser escrito como um problema de otimização, onde,

$$\text{minimizar } \frac{1}{2} \|w\|^2, \quad (3.16)$$

$$\text{condicionado a } \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon, \\ w \cdot x_i + b - y_i \leq \varepsilon. \end{cases}$$

O pressuposto para esse problema é o de que a função f existe e aproxima os pares (x_i, y_i) com precisão ε . As vezes é necessário permitir alguns erros. Assim como nos casos não lineares de máquinas de vetores de suporte, podemos acrescentar as variáveis de folga ξ_i, ξ_i^* para lidar com as condições de otimização do problema Equação 3.16, outrora impossíveis de resolver sem essas variáveis. Assim, de acordo com (VAPNIK, 1995),

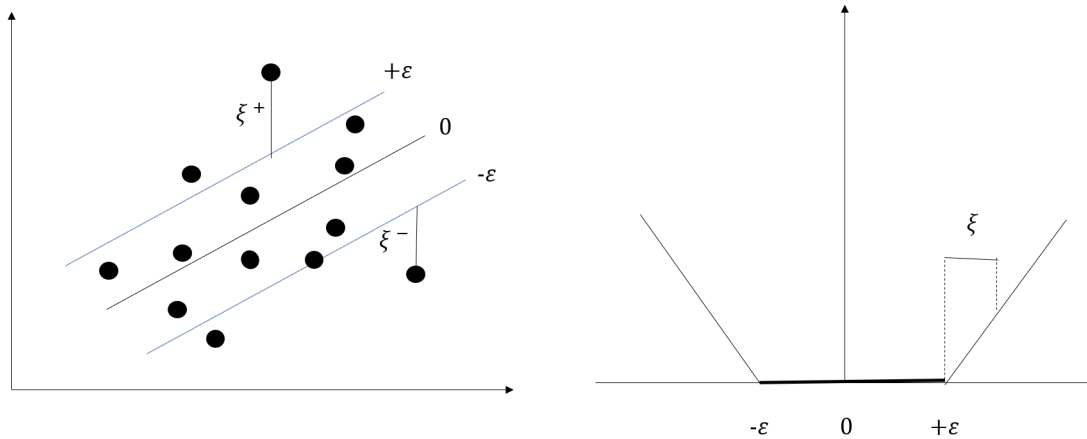
$$\text{minimizar } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (3.17)$$

$$\text{condicionado a } \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i, \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^*. \end{cases}$$

A constante $C > 0$ determina o quanto de desvio de ε é aceito. A função de perda ε -insensível $|\xi|_\varepsilon$ pode ser descrita por,

$$|\xi|_\varepsilon \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon, \\ |\xi| - \varepsilon & \text{c.c.} \end{cases}$$

Figura 1 – Representação gráfica da função de perda suave para um SVM linear



Fonte: Elaborada pelo autor.

Na maioria dos casos, o problema definido em [Equação 3.17](#) pode ser mais facilmente solucionado na forma dual. Como no caso das SVMs para classificação, monta-se um problema dual equivalente ao anterior pelo uso de uma lagrangiana, tornando nulo o resultado das derivações parciais e substituindo as expressões resultantes na equação lagrangiana inicial. Temos então ([SMOLA, 2002](#)),

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + w \cdot x_i + b) \\ & - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - y_i + w \cdot x_i - b), \end{aligned} \quad (3.18)$$

onde L é o lagrangiano, $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ são os multiplicadores de Lagrange. Assim, as variáveis dual de [Equação 3.18](#) tem que satisfazer as restrições de positividade, ou seja,

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0. \quad (3.19)$$

Segue que das derivadas parciais de L em relação as variáveis $(w, b, \xi_i, \xi_i^*,$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \quad (3.20)$$

$$\frac{\partial L}{\partial w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i = 0, \quad (3.21)$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0, \quad (3.22)$$

substituindo Equação 3.20, Equação 3.21 e Equação 3.22 resulta no problema de otimização dual, que é dado por,

$$\text{Maximizar} \begin{cases} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) x_i \cdot x_j, \\ -\varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*), \end{cases}$$

sujeito a $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$ e $\alpha_i, \alpha_i^* \in [0, C]$. (3.23)

Derivando Equação 3.23 nos já eliminamos η_i, η_i^* pela condição de Equação 3.22. assim, Equação 3.21 pode ser reescrita,

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i, \quad (3.24)$$

e assim temos que,

$$f(x) = b + \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \cdot x, \quad (3.25)$$

onde b é dado por:

$$b^* = \frac{1}{n} \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \cdot x.$$

No caso não linear $f(x_i)$ é:

$$f(x_i) = b + w_i \phi(x_i), \quad (3.26)$$

e ϕ é função de mapeamento de x para um espaço de dimensão \mathbb{R}^n . Substituindo ϕ e Equação 3.24 em Equação 3.26 temos que:

$$f(x_i) = b + \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i) \cdot \phi(x), \quad (3.27)$$

$$= b + \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i) K(x_i, x), \quad (3.28)$$

sendo K a função Kernel $\phi() \cdot \phi()$.

Assim temos que o problema de otimização é dado por (VAPNIK, 1997),

$$\min_{\alpha_i^{(*)}} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i \cdot x_j) + \varepsilon \sum_{i=1}^n (\alpha_i' - \alpha_i) - y_i \sum_{i=1}^n (\alpha_i^* - \alpha_i), \quad (3.29)$$

$$\text{Com restrições} \begin{cases} \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \\ 0 \leq (\alpha_i^*, \alpha_i) \leq C, \end{cases}$$

onde $K(x_s \cdot x_t)$ é a função Kernel de produto interno, que reduz a complexidade computacional de espaços ocultos de grande dimensão, se tornando uma parte crucial do SVR. Essa função satisfaz o teorema de Mercer então pode ser selecionado como o núcleo do SVR.

3.2 Modelos Estatísticos para Séries Temporais

Uma estratégia bastante utilizada para séries temporais é a abordagem de Box e Jenkins (BOX, 1970). Essa abordagem consiste em ajustar Modelos Auto-Regressivos Integrados de Médias Móveis (ARIMA), $ARIMA(p, d, q)$ a um conjunto de dados.

Vamos introduzir notações de operadores que serão usadas nesse capítulo: Esses operadores são:

- Operador de defasagens (*backshift*),

$$BX_t = X_{t-1}, B^m X_t = X_{t-m};$$

- Operador de translações futuro (*foward*),

$$FX_t = X_{t+1}, F^m X_t = X_{t+m};$$

- operador diferença,

$$\nabla X_t = X_t - X_{t-1}, (1 - B)X_t;$$

- Operador soma,

$$SX_t = \sum_{j=0}^{\infty} X_{t-j} = X_t + X_{t-1} + \dots = (1 + B + B^2 + \dots)X_t.$$

3.2.1 Modelos Lineares Estacionários

Modelos lineares estacionários supõem que a série temporal seja gerada através de um filtro linear (ou sistema linear), cuja entrada é ruído branco. Temos que o processo linear geral é (BOX, 1970),

$$X_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \varepsilon(B) \varepsilon_t, \quad (3.30)$$

sendo que,

$$\psi(B) = 1 + \psi_1 B + \psi_2 B + \dots, \quad (3.31)$$

é a função de transferência e μ é o parâmetro determinando o nível da série. Os erros ε_t são tais que,

$$\begin{aligned} E(\varepsilon_t) &= 0, & \forall t, \\ \text{Var}(\varepsilon_t) &= \sigma_\varepsilon^2, & \forall t, \\ E(\varepsilon_t \varepsilon_s) &= 0, & \forall s \neq t, \end{aligned}$$

onde o processo X_t é estacionário se $\sum_{j=1}^{\infty} \psi_j$ for finita ou infinita e convergente e $E(X_t) = \mu$ e a função de auto-covariância,

$$\gamma_j = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}, \quad (3.32)$$

com $\psi_0 = 1$ e função de variância,

$$\gamma_0 = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2, \quad (3.33)$$

portanto a condição de existência é $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Podemos escrever $\tilde{X}_t = X_t - \mu$ como uma soma ponderada de valores passados,

$$\tilde{X}_t = \pi_1 \tilde{X}_{t-1} + \pi_2 \tilde{X}_{t-2} + \dots + \varepsilon_t, \quad (3.34)$$

equivalentemente,

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) \tilde{X}_t = \varepsilon_t,$$

ou

$$\pi(B) \tilde{X}_t = \varepsilon_t. \quad (3.35)$$

Se substituirmos $\tilde{X}_t = \psi(B) \varepsilon_t$ segue que,

$$\pi(B) \psi(B) \varepsilon_t = \varepsilon_t,$$

de modo que,

$$\pi(B) = \psi^{-1}(B). \quad (3.36)$$

E esta relação serve para obter os pesos π_j em função dos pesos ψ_j ou vice-versa.

Proposição 1. Um processo linear será estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$; será invertível se $\pi(B)$ convergir para $|B| \leq 1$. Dado um processo estocástico, utilizando 2.4 pode-se demonstrar que a função de densidade espectral, ou espectro, de um processo linear estacionário é dada por (MORETTIN, 2006):

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} |\psi(e^{-i\lambda})|^2, \quad (3.37)$$

onde $-\pi \leq \lambda \leq \pi$, e a função ψ é,

$$\psi(e^{-i\lambda}) = 1 + \psi_1 e^{-i\lambda} + \psi_2 e^{-2i\lambda} + \psi_3 e^{-3i\lambda} + \dots,$$

ou seja, o polinômio linear geral aplicado em $e^{-i\lambda}$.

3.2.2 Processos auto-regressivos

Seja $\{\varepsilon_t\}$ um processo puramente aleatório com $E(\varepsilon_t) = 0$ e $Var(\varepsilon_t) = \sigma_\varepsilon^2$. Então $\{X_t\}$ é um processo auto-regressivo de ordem p , ou $AR(p)$, se (MORETTIN, 2006),

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t, \quad (3.38)$$

nos termos do operador de retardo temos,

$$(1 - \alpha_1 B - \dots - \alpha_p B^p) X_t = \varepsilon_t, \quad (3.39)$$

$$\phi(B) X_t = \varepsilon_t. \quad (3.40)$$

O processo $AR(p)$ é estacionário se as raízes de $\phi(B) = 0$ fora do círculo unitário. Seja $\{X_t\}$ de um processo $AR(p)$ com,

$$E(X_t) = 0, \quad Var(X_t) = \sigma_X^2, \quad Cov(X_t, X_{t-k}) = \phi(k),$$

a autocorrelação teórica do processo $AR(p)$ é calculada seguindo o seguinte procedimento,

- Multiplique o $AR(p)$ por X_{t-k} ,

$$X_t X_{t-k} = \alpha_1 X_{t-1} X_{t-k} + \dots + \alpha_p X_{t-p} X_{t-k} + \varepsilon_t X_{t-k};$$

- Tome o valor esperado,

$$\begin{aligned} E(X_t X_{t-k}) &= \gamma(k) = \alpha_1 E(X_{t-1} X_{t-k}) + \dots + \alpha_p E(X_{t-p} X_{t-k}) \\ &= \alpha_1 \gamma(k-1) + \dots + \alpha_p \gamma(k-p), k > 0; \end{aligned}$$

- Dividindo ambos os lados por σ_X^2 ,

$$= \rho(k) = \alpha_1 \rho(k-1) + \dots + \alpha_p \rho(k-p), k > 0,$$

que se fizermos $k = 1, 2, \dots, p$ obtemos,

$$\begin{aligned}\rho(1) &= \alpha_1\rho(0) + \alpha_2\rho(-1) + \dots + \alpha_p\rho(1-p), \\ \rho(2) &= \alpha_1\rho(1) + \alpha_2\rho(0) + \dots + \alpha_p\rho(2-p), \\ &\vdots \\ \rho(p) &= \alpha_1\rho(p-1) + \alpha_2\rho(p-2) + \dots + \alpha_p\rho(0),\end{aligned}$$

que são denominadas *equações de Yule-Walker*. E a variância no processo $AR(p)$ de X_t é obtida através (BOX, 1970),

- Multiplique ambos os lados por X_t no processo $AR(p)$,

$$X_t^2 = \alpha_1 X_t X_{t-1} + \dots + \alpha_p X_t X_{t-p} + X_t \varepsilon_t,$$

- Calcule a esperança,

$$\begin{aligned}E(X_t^2) &= \alpha_1 E(X_t X_{t-1}) + \dots + \alpha_p E(X_t X_{t-p}) + E(X_t \varepsilon_t) \\ &= \alpha_1 \gamma(1) + \dots + \alpha_p \gamma(p) + E(\psi(B)\varepsilon_t \varepsilon_t) \\ &= \alpha_1 \gamma(1) + \dots + \alpha_p \gamma(p) + E(\varepsilon_t^2) \\ 1 &= \alpha_1 \rho(1) + \dots + \alpha_p \rho(p) + \sigma_\varepsilon^2 / \sigma_X^2 \\ \sigma_X^2 &= \frac{\sigma_\varepsilon^2}{1 - \alpha_1 \rho(1) - \dots - \alpha_p \rho(p)},\end{aligned}$$

como não existe correlação direta entre X_t e $X_{t-p-1}, X_{t-p-2}, \dots$, substituindo $k = p+1, p+2, \dots$ nas equações de Yule-Walker temos que os coeficientes de correlação parciais serão nulos para todo $k > p$, e esse fato sugerido é usado como ferramenta para determinar a ordem de p no processo autoregressivo das séries temporais.

3.2.3 Processo de Médias móveis

Se considerarmos um processo linear, seja $\{\varepsilon_t\}$ um processo discreto puramente aleatório com $E(\varepsilon_t) = 0$ e $Var(\varepsilon_t) = \sigma_\varepsilon^2$. $\{X_t\}$ é chamado de processo de médias móveis de ordem q , ou $MA(q)$, se (MORETTIN, 2006),

$$X_t = \varepsilon_t + \beta_{1\varepsilon_{t-1}} + \dots + \beta_{q\varepsilon_{t-q}}, \quad (3.41)$$

sendo $\beta_i \in \mathbb{R}$, $i = 1, \dots, q$. A média e a variância nesse processo ficam,

$$E(X_t) = E(\varepsilon_t) + \sum_{j=1}^q \beta_j E(\varepsilon_{t-j}) = 0,$$

$$\text{Var}(X_t) = \text{Var}(\varepsilon_t) + \sum_{j=1}^q \beta_j^2 \text{Var}(\varepsilon_{t-j}) = (1 + \beta_1^2 + \dots + \beta_q^2) \sigma_\varepsilon^2.$$

Verifique a estrutura de auto-covariância,

$$\begin{aligned} \gamma(k) &= \text{Cov}(X_t, X_{t+k}) = \\ &= \text{Cov}(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \varepsilon_{t+k} + \beta_1 \varepsilon_{t+k-1} + \dots + \beta_q \varepsilon_{t+k-q}), \\ &= \begin{cases} 0, & k > q, \\ \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k}, & k = 0, \dots, q, \\ \gamma(-k), & k < 0, \end{cases} \end{aligned} \quad (3.42)$$

com $\beta_0 = 1$. $\gamma(k)$ não depende de t pois o processo é estacionário para todos β_1, \dots, β_q e média e variância são constantes. Se ε_t for normalmente distribuídos temos que X_t também será e portanto o processo será estritamente estacionário. A função de autocorrelação é dada por,

$$\rho_j = \begin{cases} \frac{\sum_{j=0}^{q-k} \beta_j \beta_{j+k}}{\sum_{j=0}^q \beta_j^2}, & k = 1, \dots, q, \\ 0, & j > q. \end{cases}$$

3.2.4 Modelos Auto Regressivos e de Médias móveis

Combinando os processos AR e MA pode-se obter uma representação adequada com um número menor de parâmetros. O modelo $ARMA(p, q)$ é dado por (MORETTIN, 2006),

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \quad (3.43)$$

sendo $\{\varepsilon_t\}$ um processo puramente aleatório com média zero e variância σ_ε^2 .

O processo é estacionário se os valores $\alpha_1, \dots, \alpha_p$ são tais que as raízes de $\phi(B) = 0$ estão fora do círculo unitário. O processo é inversível se os valores de β_1, \dots, β_q são tais que as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

3.2.5 Modelos ARIMA

Quando a série é não estacionária, um modelo ARMA no qual X_t é substituído pela sua d -ésima diferença $\nabla^d X_t$ é capaz de descrever alguns tipos de séries não estacionárias (EHLERS, 2009).

Denotamos a série diferenciada por,

$$W_t = \nabla^d X_t = (1 - B)^d X_t, \quad (3.44)$$

o processo autoregressivo integrado médias móveis é dado por,

$$W_t = \alpha_1 W_{t-1} + \dots + \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}. \quad (3.45)$$

Denotamos o modelo $ARIMA(p, d, q)$ como (MORETTIN, 2006),

$$\phi(B)(1-B)^d X_t = \theta(B)e_t, \quad (3.46)$$

onde $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ e a série e_t é Gaussiana $N(0, \delta_e^2)$.

O modelo para X_t é claramente não estacionário já que o polinômio autoregressivo $\phi(B)(1-B)^d$ tem exatamente d raízes sobre o círculo unitário, ou d raízes unitárias. O critério de informação é uma forma de discriminar possíveis modelos competidores, que levam em conta não apenas a qualidade do ajuste mas também penalizam a inclusão de parâmetros extras. Tem-se como regra básica utilizar o modelo cujo critério de informação calculado seja mínimo.

Suponha que existem k diferentes modelos M_1, \dots, M_k onde para cada modelo existe um vetor de parâmetros $\theta_i \in \mathbb{R}^{n_i}$, $i = 1, \dots, k$, para cada modelo existe uma função de verossimilhança $p(x|\theta_i, M_i)$, para selecionar entre esse modelos temos os seguintes critérios de informação

3.2.5.1 Critério de informação

O Critério de informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC) tem a forma (BOX, 1970),

$$AIC = -2 \log p(x|\hat{\theta}_i, M_i) + 2m, \quad (3.47)$$

$$BIC = -2 \log p(x|\hat{\theta}_i, M_i) + m \log n, \quad (3.48)$$

onde $-2 \log p(x|\hat{\theta}_i, M_i)$ é o logaritmo da verossimilhança maximizada, sendo m o número de parâmetros no modelo. Assim, busca-se modelos parcimoniosos, pois contém um número pequeno de parâmetros e as previsões obtidas são bastante precisas, comparando-se favoravelmente a outros métodos de previsão.

3.3 Aplicação das técnicas

Nesse capítulo descrevemos os métodos que serão utilizados nesse trabalho, sendo aplicados uma técnica de aprendizado de máquina, o SVR, e um modelo estatístico clássico, ARMA. A seguir vamos apresentar os resultados na aplicação dessas técnicas, assim como uma descrição dos dados que estamos utilizando.

EXPERIMENTO E RESULTADOS

Nesse capítulo apresentamos o tratamento e uma análise descritiva dos dados, para depois aplicarmos as técnicas selecionadas em dados simulados pela série de Lorenz, e assim aplicando as técnicas nos dados reais de fechamento da moeda ETH.

A seguir apresentamos os resultados desse estudo.

4.1 Análise Exploratória

Ethereum é uma blockchain pública, de código aberto e sistema operacional com a funcionalidade de contratos inteligentes. ETH é um “*token*” gerado pela blockchain para recompensar os participantes no nós de mineração que performaram computação. Esse *token* tem um valor monetário, que pode ser transacionado e transferido. No mercado especulativo de criptomoedas, o ETH vem sendo comercializado desde maio de 2016. Assim, coletamos os valores da cotação, em dólares, do Ether, por hora, desde a abertura do seu mercado de *tokens*. Abaixo apresentamos uma análise descritiva dos dados de cotação do fechamento do valor da criptomoeda ETH em dólar no período de 09/05/2016 13:00 a 17/09/2018 23:00 coletados por hora. As variáveis presentes no banco de dados são:

- Unix.Timestamp – Código de data
- Date – Data no formato dd/MM/yyyy hh:mm
- Simbolo – Simbolo da transação (ETHUSD – ETH para Dólar)
- Fechamento – Valor de Fechamento em dólares

Os dados foram coletados do site de câmbio internacional, *Coindesk*, através do site *CriptoDownloads* ([CRYPTODATA, 2018](#)) em um arquivo .csv com as variáveis citadas acima.

Carregamos os dados no software de código aberto R, onde fizemos tratamento e a análise dos dados.

Tabela 1 – Estatísticas descritivas ETH por ano

Fechamento (média +/- DP (Min - Max))	
Anual	
2016	11,52 +/- 2,10 (6,10 - 21,09)
2017	220,56 +/-182,76 (8,15 - 860,99)
2018	606,97 +/-259,80 (169,79 - 1417,54)

Fonte: Dados da pesquisa.

A [Tabela 1](#) nos mostra que, como esperado, a abertura do comércio do *token* Ether em 2016, o mínimo comercializado foi de U\$0 USD, ou seja, ele foi comercializado por uma parcela menor do que um centavos de dólar, mas durante os 6-7 meses depois de sua abertura, a moeda se valorizou e fechou com o valor médio de U\$11,52 e máxima de alta em U\$25,05 dólares.

Tabela 2 – Estatísticas descritivas ETH por mês

Fechamento (média +/- DP (Min - Max))		
	2016	2017
Janeiro	-	10,20 +/- 0,69 (8,15 - 11,84)
Fevereiro	-	12,29 +/-1,26 (10,50 - 16,29)
Marco	-	33,89 +/-13,32 (15,87 - 53,88)
Abril	-	50,03 +/-7,55 (41,42 - 79,71)
Mai	12,04 +/- 1,52 (9,36 - 14,75)	122,16 +/-45,51 (75,45 - 233,01)
Junho	14,38 +/- 1,99 (10,50 - 21,09)	300,9 +/-50,08 (211,05 - 394,90)
Julho	12,02 +/- 1,10 (9,73 - 15,50)	220,61 +/-32,70 (137,00 - 291,98)
Agosto	11,24 +/- 0,54 (7,81 - 12,68)	298,17 +/-43,01 (205,70 - 388,58)
Setembro	12,51 +/- 0,71 (10,50 - 14,38)	297,05 +/-33,81 (203,44 - 394,63)
Outubro	12,16 +/- 0,72 (9,88 - 13,70)	305,32 +/-14,10 (277,21 - 348,90)
Novembro	10,09 +/-0,76 (8,00 - 11,95)	354,88 +/-63,67 (279,43 - 503,82)
Dezembro	7,90 +/-0,49 (6,10 - 8,75)	624,84 +/-135,19 (405,79 - 860,99)

Fonte: Dados da pesquisa.

A [Tabela 2](#) mostra que entre os meses maio e agosto de 2016, o Ether ficou estável, chegando ao valor médio de fechamento de U\$21,09 e alta de U\$25,09 em junho, e mínimo de U\$9,36 de fechamento em maio e baixa de U\$11,19 em agosto. Observamos uma queda nos meses de setembro a dezembro em 2016, chegando ao valor médio de fechamento de U\$21,09 em junho, e mínimo de U\$6,10 em dezembro. Com relação a maio, o fechamento médio em dezembro foi menor do que o de Maio, apresentando queda em novembro, e estando estável na média de maio a outubro. Em 2017 de janeiro a junho, o *token* saiu de uma média de 10 dólares para um média de fechamento de U\$300 em junho. O crescimento acentuado começa a partir da média em março, subindo mais acentuadamente a partir de maio, onde observa-se um crescimento médio no fechamento 2800% maior do que no mês de janeiro. De julho a dezembro

Tabela 3 – Estatísticas descritivas ETH por mês no ano 2018

Fechamento (média +/- DP (Min - Max))	
	2018
Janeiro	1083,99 +/- 144,74 (722,11 - 1417,54)
Fevereiro	872.32 +/-78,38 (573,20 - 1144,87)
Marco	629.69 +/-149,11 (368,68 - 876,98)
Abril	516.02 +/-112,80 (361,15 - 708,75)
Mai	676.36 +/-75,07 (511,37 - 827,00)
Junho	519.78 +/-61,33 (407,34 - 632,78)
Julho	461.71 +/-16,75 (424,56 - 509,00)
Agosto	322.85 +/-53,49 (253,82 - 431,53)
Setembro	229.44 +/-38,80 (169,79 - 301,59)
Outubro	-
Novembro	-
Dezembro	-

Fonte: Dados da pesquisa.

no mesmo ano, o *token* saiu de uma média de U\$220 dólares para um média de fechamento de U\$860,99. Observamos também uma variância e desvio padrão altos nos meses desses ano. O crescimento acentuado começa a partir da média em março, subindo mais acentuadamente a partir de maio, se mantendo com uma média em volta de U\$200 - U\$300 até novembro, onde observa-se um crescimento médio no fechamento 76% maior do que no mês anterior.

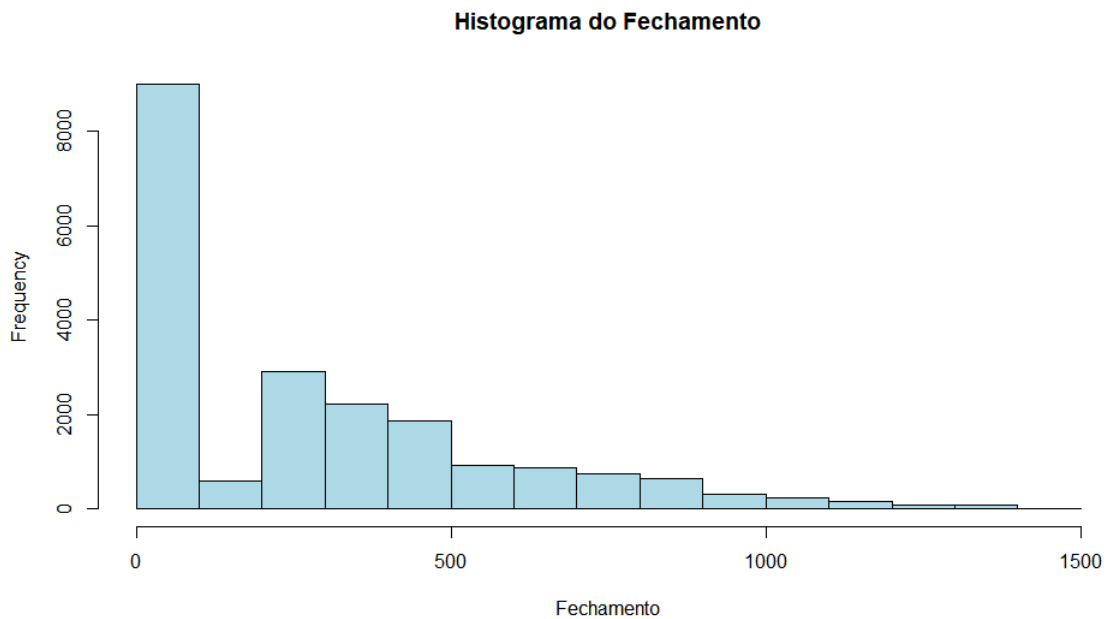
Em 2018 A [Tabela 3](#) uma média de fechamento de janeiro em US1084, ou seja, um crescimento de 75% em relação ao mês anterior, caindo durante o ano até maio em 37% do valor de fechamento em janeiro, para U\$676 dólares. Observamos também em janeiro um pico de U\$1420 dólares, mostrando que a maior alta da série foi nesse mês. Com um risco muito alto, mas rápidos retornos, nesse mês muitas pessoas resolveram adquirir criptomoedas com o objetivo de ter retornos rápidos, no mês de fevereiro, a resposta do mercado foi de uma alta queda, caracterizando como uma bolha que surgiu durante esse período, dado que houve uma queda acentuada durante os meses subsequentes de junho começando em U\$519,79, caindo durante o ano até setembro em 78% do valor de fechamento em janeiro, para U\$229 dólares. Observamos também em janeiro um pico de U\$1417,54 dólares, mostrando que a maior alta da série foi nesse mês.

Figura 2 – Cotação por hora



A [Figura 2](#) mostra o comportamento hora a hora do ETHER comercializado. Podemos observar o pico em 2018, e a queda que foi subsequente durante o ano de 2018.

Figura 3 – Histograma do Fechamento



A [Figura 3](#) apresenta histograma de fechamento mostrando a frequência dos valores durante todo o período. Observamos uma assimetria na calda a esquerda na distribuição dos valores.

4.2 Simulação de dados e comparação de modelos

Para avaliar a performance das técnicas que serão utilizadas nesse estudo, dado uma variável $y_t, t = 1, \dots, T$ foi gerada aleatoriamente através de um sistema não linear de Lorenz com os parâmetros baseados em (CHEN KIHU JEONG, 2008),

$$\frac{\partial y}{\partial t} = \sigma(x - y), \quad (4.1)$$

$$\frac{\partial x}{\partial t} = -yz + \rho y - x, \quad (4.2)$$

$$\frac{\partial z}{\partial t} = -yx + \beta z, \quad (4.3)$$

$$(4.4)$$

onde $\sigma = 16$, $\beta = 45,92$ e $\rho = 4$. Um ruído t-student foi incluído na simulação contínua da série y_t . Na simulação foi gerada uma amostra $T = 1000$ com 200 replicações. Para avaliarmos a qualidade do modelo vamos utilizar duas medidas padrões de performance para o modelo, a primeira, raiz quadrada do erro quadrático médio (Root Mean Squared Error (RMSE) e o erro médio absoluto Mean Absolute Error (MAE) (MAATALLAH AJIT ACHUTHAN, 2015).

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t+h) - \hat{y}(t+h|t))^2},$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |(y(t+h) - \hat{y}(t+h|t))|,$$

onde utilizaremos o *kernel* laplaciano $k(x, x') = \exp(-\sigma||x - x'||)$ para aplicar o ε -SVR, com hiperparâmetro $C = 1$. As previsões serão feitas um passo a frente para o modelo ARMA, ε -SVR e SVR-ARMA, onde propomos pesos $SVR_{rmse}/(SVR_{rmse} + ARMA_{rmse})$ e $ARMA_{rmse}/(SVR_{rmse} + ARMA_{rmse})$ para os valores preditos dos modelos SVR e ARMA respectivamente.

A série foi dividida em uma amostra de treinamento e de teste, onde temos 777 observações de treinamento, e 24 observações para teste do modelo. As previsões foram realizadas sempre um passo a frente, atualizando o modelo para cada nova informação dos dados de teste.

Os resultados gerados pelos dados simulados são apresentados a seguir. A Tabela 4 mostra que o SVR ajustado apresenta melhor resultado nos dados de treinamento ($RMSE = 0,8872$ e $MAE = 0,7879$) e resultados. Quando ponderamos os resultados de previsão de ambos os modelos (SVR-ARMA), podemos observar uma melhora na previsão ($RMSE = 2,2302$ e $MAE = 1,7783$ em relação aos resultados obtidos pelos modelos individualmente. A Figura 4 apresenta o gráfico com os resultados da simulação. Podemos observar que os valores ajustados estão bem próximos dos valores reais. E os valores preditos apresentam um pequena distância

dos valores reais. A [Figura 5](#) apresenta o comportamento dos 3 modelos em relação a previsão de 24 passos a frente na série.

Podemos observar que a ponderação dos resultados dos modelos, resultando no que chamamos de modelo SVR-ARMA, trazem resultados mais precisos do que o modelo ARMA e SVR individualmente. Assim, vamos a seguir aplicar as técnicas aqui mencionadas em dados reais.

Tabela 4 – Performance dos Modelos - Dados simulados

Modelos	Treinamento		Teste	
	RMSE	MAE	RMSE	MAE
SVR	0,8872	0,7879	3,662	2,5095
ARMA(2,3)	1,5575	1,2164	2,2344	1,8528
SVR-ARMA	-	-	2,2302	1,7783

Fonte: Dados da pesquisa.

Figura 4 – Resultados da Simulação

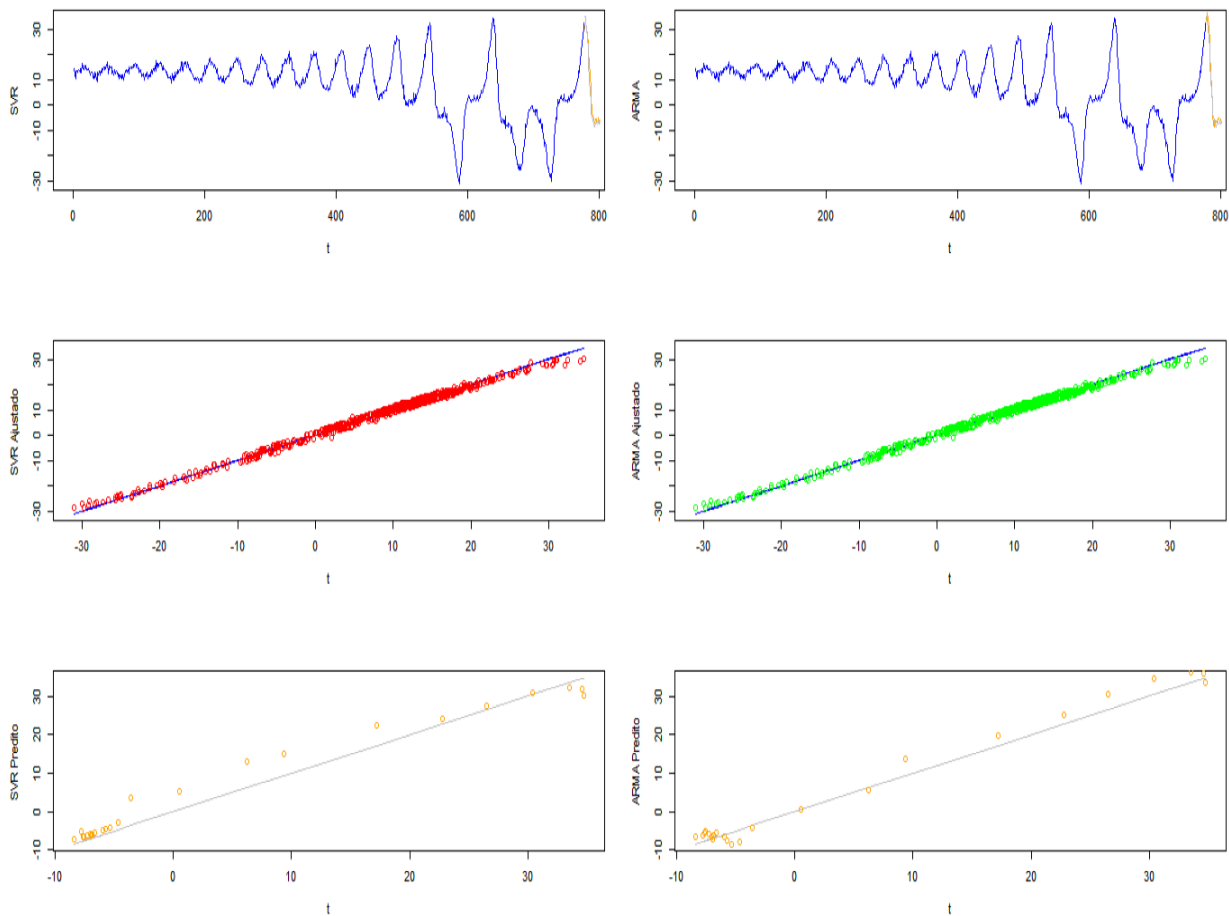
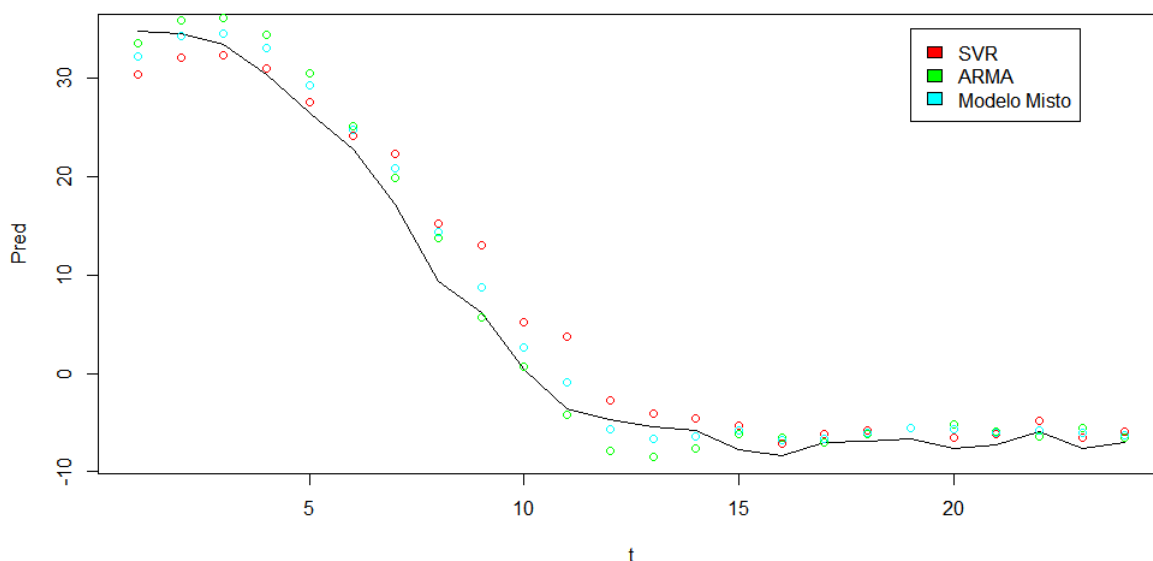


Figura 5 – Atual vs Preditos (SVR, ARMA e SVR-ARMA)



4.3 Aplicação dos métodos para predição da Ethereum

A seguir apresentamos e comparamos os resultados das técnicas mencionadas a série de cotações da ETH. A modelagem foi realizada utilizando o software R em um computador com as seguintes especificações (CPU Core i5-7300 2,5GHz, 8gb RAM, 64 bits OS).

Como o preço de ativos são baseado em retornos, a [Figura 6](#) mostra que os dados são não estacionários, para esse tipo de série aplicamos uma transformação chamada *log-retornos* ([MORETTIN, 2006](#)) nos dados, que consiste em,

$$y_t = \log \frac{P_t}{P_{t-1}} = \log(P_t) - \log(P_{t-1}),$$

onde y é o *log-retorno*, e P o valor de fechamento no tempo t , $t \in 1, \dots, n$.

A série contém 20.675 observações, que foram divididas em 20.651 observações de treinamento e 24 para teste, onde vamos realizar a predição um passo a frente para todos os casos. A [Figura 7](#) mostra a série transformada e seu gráfico de auto correlação. Aplicamos o modelo ARMA(p,q), o modelo usando vetor de suporte foi utilizado o *kernel* Laplaciano.

O método aplicado pelo SVR apresentou $RMSE = 0,96$ e $MAE = 0,8266$ demonstrando melhor performance do que o modelo ARMA(2,3) com $RMSE = 1,5572$ e $MAE = 1,216$.

Figura 6 – Série original da Ethereum e sua função de auto-correlação

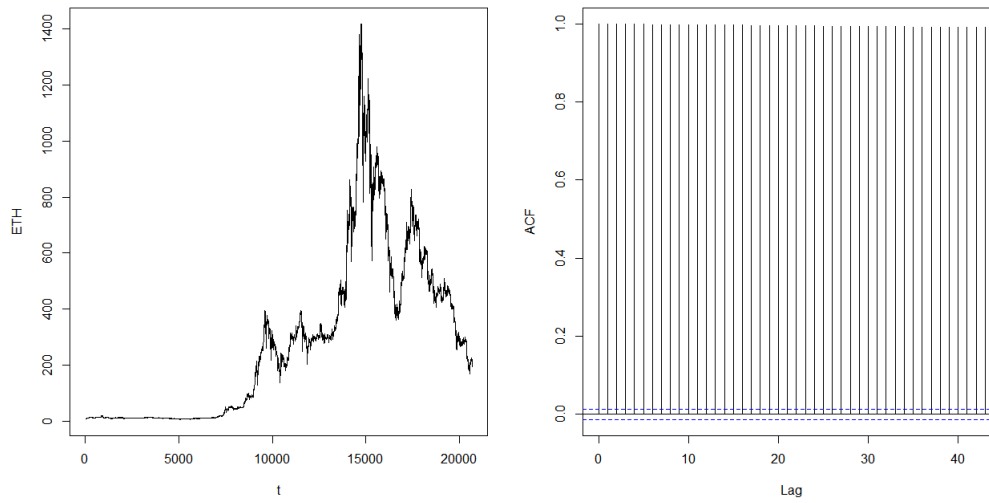


Figura 7 – Série original da Ethereum e sua função de auto-correlação

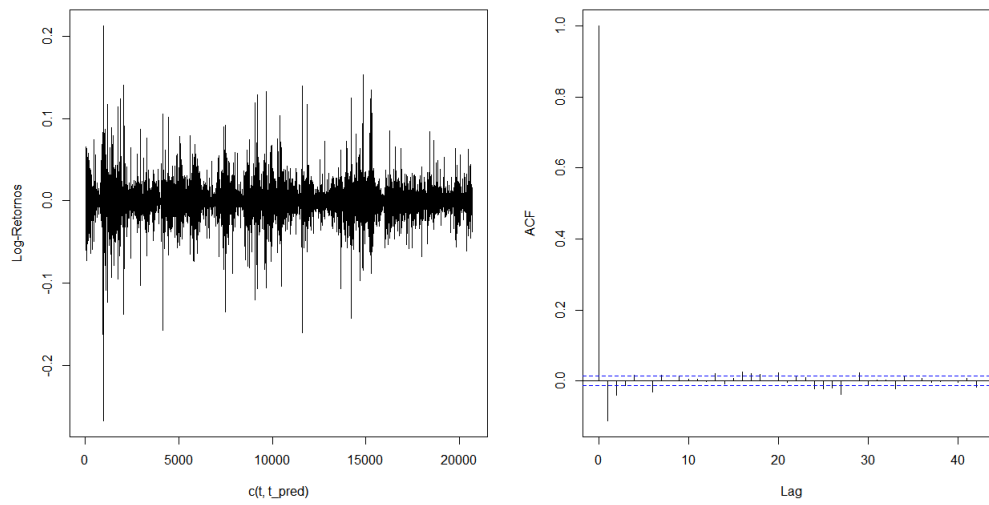


Tabela 5 – Performance dos Modelos - Dados reais

Modelos	Treinamento		Teste	
	RMSE	MAE	RMSE	MAE
SVR	0,9696	0,8266	3,1572	2,4205
ARMA(2,3)	1,5572	1,216	2,2344	1,8528
SVR-ARMA	-	-	2,0834	1,7569

Fonte: Dados da pesquisa.

Figura 8 – Série original da Ethereum e Valores preditos pelos modelos

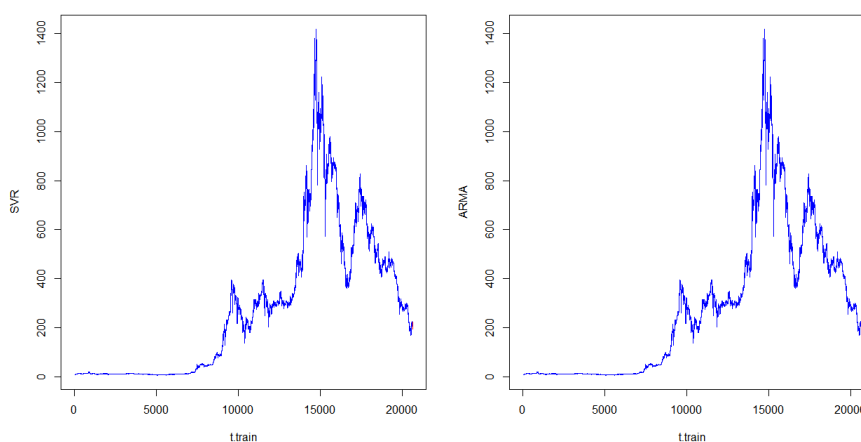
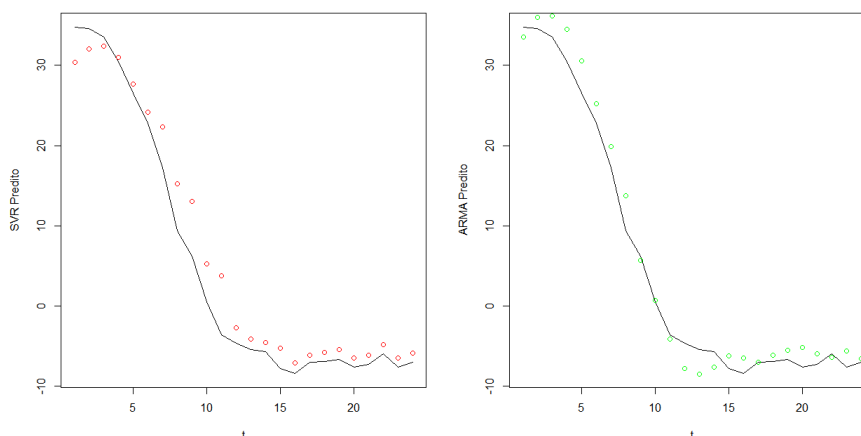


Figura 9 – Amostra Teste e Valores preditos pelos modelos



Seguimos combinando a predição de ambos os modelos (SVR-ARMA) ponderados pela pesos $SVR_{rmse}/(SVR_{rmse} + ARMA_{rmse})$ e $ARMA_{rmse}/(SVR_{rmse} + ARMA_{rmse})$, e assim calculamos o RMSE e MAE dos modelos. Quando vamos prever um passo a frente, o modelo $ARMA(2,3)$ apresentou $RMSE = 2,23$ e $MAE = 1,85$, indicando uma performance ao prever melhor do que o SVR ($RMSE = 3,15$ e $MAE = 2,42$).

A [Tabela 5](#) apresenta os resultados de qualidade do ajuste, a combinação dos modelos

Figura 10 – Amostra Teste e Valores preditos pelos modelos

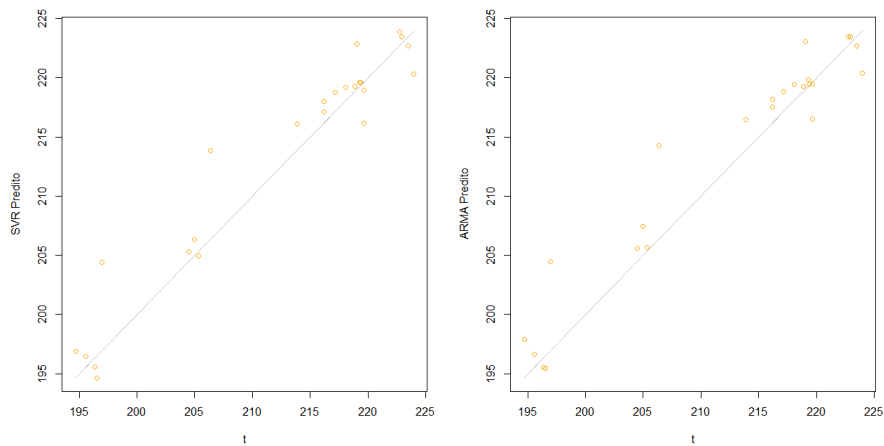
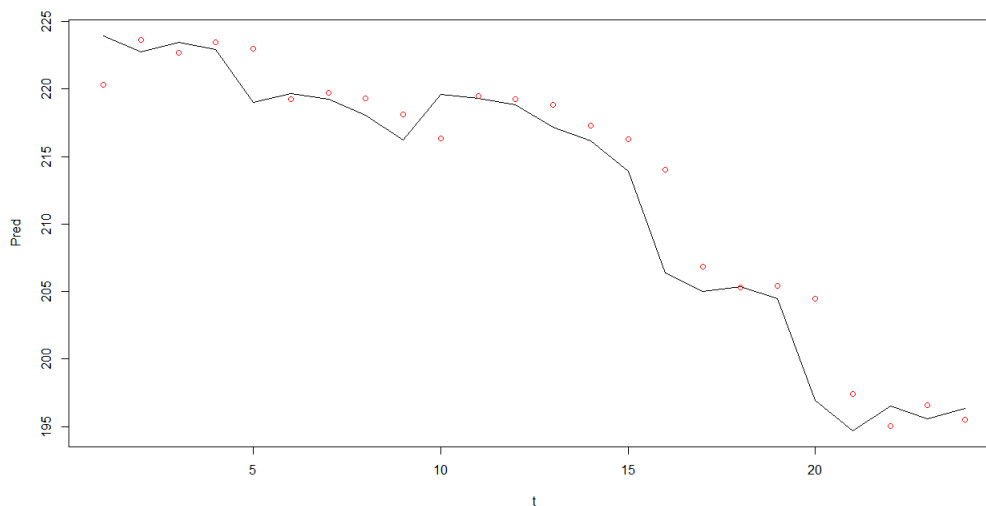


Figura 11 – Ajuste ponderado SVM-ARMA



SVR-ARMA apresenta uma melhora no ajuste, tornando o modelo mais preciso.

As figuras [Figura 9](#) e [Figura 10](#) mostram o comportamento do ajuste em relação a amostra de teste. Podemos observar que os modelos apresentam um bom ajuste ao acompanhar o sentido da série, podemos observar que o SVR e o modelo ARMA(2,3) estão ajustando bem nesse quesito. A [Figura 10](#) mostra que os pontos preditos estão próximos em relação a série original. Encontramos pontos que apresentam uma distância maior em relação a série original, mas o grau de precisão apresentado pelos indicadores mostra que o nosso modelo apresenta uma boa performance de predição da série.

Ao combinar os modelos conseguimos aumentar um pouco mais a precisão das predições, alcançando um $RMSE = 2,08$ e $MAE = 1,75$, e na [Figura 11](#) podemos ver que o modelo se ajusta melhor aos dados preditos. Os pesos estimados para o modelo são para $SVR = 0,6162$ e para os resultados preditos para o modelo $ARMA = 0,3837$.

4.4 Comentários sobre a Aplicação

Nesse capítulo descrevemos o fechamento por hora do criptoativo ETH no período de 05/2016 a 09/2018, com grande volatilidade durante esse período, encontramos uma amplitude de U\$10 a U\$1400 durante o período. Testamos a aplicação do modelo ARMA e do SVR com o Kernel laplaciano em uma série gerada aleatoriamente por Lorenz, que apresentou um bom ajuste, com $RMSE_{SVR} = 0,95$ e $RMSE_{ARMA} = 1,55$ para os modelos ajustados. Seguimos aplicando as técnicas nos dados reais, obtendo 2 modelos, um SVR e outro ARMA, assim como uma combinação da predição desses modelos que chamamos de SVR-ARMA.

A combinação desses dois modelos apresentou uma melhor precisão em relação a previsão dos modelos sozinhos com $RMSE_{SVR-ARMA} = 2,08$. A seguir concluímos esse trabalho com os comentários e considerações finais.

CONSIDERAÇÕES FINAIS SOBRE O TRABALHO

Nesse estudo discutimos a tecnologia de *Blockchain* introduzida por (NAKAMOTO, 2008), que introduziu uma nova forma de pagamento e uma nova moeda, a Bitcoin, eliminando a necessidade de terceiros para verificar o pagamento digital. A *Blockchain* é uma tecnologia presente em diferentes serviços, como contratos inteligentes, transferências de moeda internacionais, armazenamento em nuvem e etc. A Ethereum é uma das precursoras dessa tecnologia, atuando com uma *Blockchain* própria de contratos inteligentes, e gerando a criptomoeda ETH.

Apresentamos a aplicação de máquinas de vetores de suporte e um modelo estatístico clássico em uma dados de cotação, por hora, da criptomoeda Ethereum.

Inicialmente testamos a metodologia a ser aplicada em uma simulação de dados gerada pela série de Lorenz, onde aplicamos o modelo ARMA e a máquina de vetor de suporte em regressão utilizando o *kernel* Laplaciano. Durante o treinamento do modelo observamos um melhor desempenho da máquina de vetor de suporte em relação ao modelo clássico estatístico ARMA(2,3), indicadores de RMSE e MAE apresentaram melhores resultados para a máquina de vetor de suporte. Seguimos fazendo uma combinação dos modelos, SVR-ARMA aplicando pesos ($SVR = 0,61$ e $ARMA = 0,39$) para os valores preditos, assim aumentando a precisão do modelo em relação a amostra de teste. Validando assim a aplicação das técnicas em dados reais. Devido a natureza de uma série financeira e não estacionariedade desse tipo de série, foi realizada a transformação dos dados em uma forma mais apropriada para o ajuste do modelo. A transformação utilizada foi a *log-retorno*, que resultou em uma série temporal estacionária, e assim aplicamos as metodologias propostas.

Para testar a acurácia do modelo, foram realizadas predições um passo a frente em todos os casos, e a precisão foi medida durante todo o estudo.

Ambos os modelos conseguem capturar com sucesso a direção da série, um aspecto

importante para séries financeiras. Nas simulações, o modelo SVR conseguiu com mais eficiência e acurácia gerar valores preditos para a série, assim como o ajuste a série simulada original.

Aplicando as técnicas na série de dados reais o modelo SVR com o *kernel* Laplaciano apresentou um ajuste a série original mais preciso do que o modelo $ARMA(2,3)$, mas a acurácia ao predizer um valor a frente não foi significativamente melhor que o modelo $ARMA(2,3)$ estimado.

Combinando então os resultados dos dois modelos utilizando os pesos, o modelo SVR-ARMA apresentou acurácia nos valores preditos maior do que os outros dois modelos.

Esse trabalho pode ser continuado de diversas formas, devido à riqueza de informações relacionadas a esse tipo de ativos, aplicar outros tipos de modelagens, comparando com diferentes tipos de *kernels*, acrescentando mais entradas e diferentes variáveis, combinando outras técnicas clássicas estatísticas com máquinas de vetores de suporte, buscar predições k passos a frente e explorar técnicas em busca de obter melhor performance de resultados.

REFERÊNCIAS

- BHOSALE, S. M. J. Volatility of selectcrypto-currencies: A comparison of bitcoin, ethereum and litecoin. **SCMS Pune**, v. 6, p. 132–141, 2018. Citado nas páginas 23, 24 e 27.
- BOX, G. M. J. G. E. P. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden Day, 1970. Citado nas páginas 40, 41, 43 e 45.
- BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. **Knowledge Discovery and Data Mining**, p. 1–43, 1998. Citado na página 35.
- CHEN KIHONG, W. K. H. S. Recurrent support vector regression for a nonlinear arma model with applications to forecasting financial returns. **Humboldt-Universität zu Berlin**, SFB 649, 2008. Citado nas páginas 28 e 51.
- _____. Forecasting volatility with support vector machine-based garch model. **Jornal of Forecasting**, v. 29, n. 4, p. 406–433, 2010. Citado nas páginas 28 e 35.
- CRYPTODATA. **Cryptodata**. 2018. Disponível em: <<https://www.cryptodatadownload.com/data>>. Acesso em: 13/10/2018. Citado na página 47.
- EHLERS, R. S. **Análise de Séries Temporais**. 2009. Disponível em: <<http://conteudo.icmc.usp.br/pessoas/ehlers/stemp/stemp.pdf>>. Citado na página 44.
- FACELI ANA CAROLINA LORENA, J. G. A. C. P. d. L. F. d. C. K. **Inteligência artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro, RJ: LTC Grupo Editorial Nacional, 2011. Citado na página 36.
- HAYKIN, S. **Neural Networks: a comprehensive foundations**. New Jersey: Prentice Hall, 1999. Citado na página 36.
- INVESTOPEDIA.COM. **investopedia.com**. 2018. Disponível em: <<https://www.investopedia.com/technical-analysis-4689657>>. Acesso em: 13/10/2018. Citado nas páginas 29 e 30.
- KPMG. The pulse of fintech 2018. **KPMG**, 2018. Citado na página 27.
- MAATALLAH AJIT ACHUTHAN, K. J. P. M. O. A. Applied energy. **Elsevier**, p. 191–197, 2015. Citado na página 51.
- MAN-CHUNG WONG CHI-CHEONG, L. C.-C. C. Financial time series forecasting by neural network using conjugate gradient learning algorithm and multiple linear regression weight initialization. **The Hong Kong Polytechnic University**, 2018. Citado na página 28.
- MITCHELL, M. J. e T. Machine learning:trends, perspectives, and prospects. **Science Mag**, Science 349, n. 255, p. 37–48, 2015. Citado na página 31.
- MITCHELL, T. **Machine Learning**. New York: Mcgraw-Hill, 1997. Citado na página 31.
- MORETTIN, C. T. P. A. **Análise de séries temporais**. [S.l.: s.n.], 2006. Citado nas páginas 28, 29, 42, 43, 44, 45 e 53.

- MULLER K. R., M. S.-R. G. T.-K. e. S. B. **An introduction to kernel-based learning algorithms**. [S.l.]: IEEE Transactions on Neural Networks, 2001. Citado na página 32.
- NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. Outubro 2008. Disponível em: <<https://bitcoin.org/bitcoin.pdf>>. Acesso em: 13/10/2018. Citado nas páginas 23, 27 e 59.
- PENG, Y. **Support Vector Regression aplicado à previsão de taxas de cambio**. Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2016. Citado nas páginas 27 e 28.
- RAYMAKERS. Cryptocurrency bitcoin: Disruption, challenges and opportunities. **Journal of Payments Strategy Systems**, p. 1–40, 2014. Citado na página 27.
- ROSILLO J. GINER, D. D. I. f.-R. P. R. Trading system based on support vector machines in the sp500 index. **University of Oviedo, University of La Laguna**, 2018. Citado na página 31.
- SMOLA, B. S. A. **A tutorial on Support Vector Regression**. NC2-TR-1998-030, 1998. Citado nas páginas 34, 35, 36 e 37.
- SMOLA, B. S. A. J. **Learning with Kernels**. Cambridge, MA: The MIT Press, 2002. Citado nas páginas 32, 34, 37 e 38.
- SWAN, M. Blockchain: A blueprint for a new economy. **Journal of Information Systems**, p. 155–167, 2015. Citado na página 27.
- VAPNIK, V. **The Nature of Statistical Learning Theory**. New York: Springer, 1995. Citado nas páginas 31, 33 e 37.
- _____. **Statistical Learning Theory**. New York: Wiley, 1997. Citado nas páginas 33 e 40.
- WANG, J. K. J. Predicting stock price trend using macd optimized by historical volatility. **Mathematical Problems in Engineering**, p. 12, 2018. Citado na página 30.
- WILDER, J. J. W. **New Concepts in Technical Trading Systems**. Greensboro, N.C: Hunter Publishing Company, 1978. Citado na página 30.

