



ANÁLISE E MODELAÇÃO DO MERCADO DE AÇÕES

MÁRIO ROBERTO DOS REIS GOMES

setembro de 2022

ANÁLISE E MODELAÇÃO DO MERCADO DE AÇÕES

Mário Roberto Dos Reis Gomes

2022

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

ANÁLISE E MODELAÇÃO DO MERCADO DE AÇÕES

Mário Roberto Dos Reis Gomes

Estudante n.º 1171413

Dissertação apresentada ao Instituto Superior de Engenharia do Porto para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia e Gestão Industrial, realizada sob a orientação do Doutor Carlos Manuel Abreu Gomes Ferreira.

2022

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

AGRADECIMENTOS

Sendo que a presente dissertação representa para mim o culminar de mais uma etapa, é importante expressar os meus sinceros agradecimentos a todas as pessoas que contribuíram, direta ou indiretamente, para que a elaboração deste projeto fosse possível.

Contudo, quero agradecer particularmente:

Ao professor Carlos Manuel Abreu Gomes Ferreira por toda a disponibilidade e apoio demonstrados ao longo do projeto.

À minha família e amigos pela compreensão, paciência e, acima de tudo apoio demonstrado tanto ao longo de todo o trabalho como ao longo de todas as etapas na vida, em especial nas alturas que mais precisei do seu apoio.

A todos, o meu sincero OBRIGADO.

página propositadamente em branco

RESUMO

A presente dissertação, realizada no âmbito do Mestrado em Engenharia e Gestão Industrial (MEGI), é o resultado de um projeto individual, com o objetivo de prever o mercado de ações para uma base temporal de curto prazo, utilizando para isso modelos de *machine learning*. Foram consideradas para este estudo, apenas ações presentes no *SP500* de 2013 a 2018 sensivelmente.

Embora a previsão do mercado de ações seja um desafio que tem constantemente motivado a comunidade científica a procurar respostas desde a sua existência, apenas nos últimos anos com a elevada capacidade computacional existente, é que se tem verificado a existência de alguns modelos capazes de prever o mercado de ações de forma lucrativa (Rouf et al., 2021). De realçar que mesmo estes casos de sucesso ainda mostram bastantes fragilidades ao nível da consistência (Rouf et al., 2021). O que se pretende com este projeto é analisar um conjunto de dados macroeconómicos no sentido de concluir se os fatores presentes permitem alcançar melhores resultados de previsão que um modelo de persistência, que considera como valor de previsão, o preço de fecho do dia anterior.

Para a realização deste projeto seguiu-se a metodologia CRISP-DM, muito utilizada em projetos de *data mining*. Através desta metodologia foi possível definir objetivos claros de trabalho, descrever e analisar o *dataset* utilizado, preparar os dados tendo em conta os modelos utilizados, submeter o *dataset* aos três modelos utilizados, nomeadamente o *Linear Regression*, o *Random Forest Regression* e o *MultiLayer Perceptron*, e por fim, tecer as devidas conclusões face aos resultados obtidos.

Na fase da aplicação dos modelos também se teve em consideração a realização de análises de sensibilidade face aos valores de parâmetros a utilizar, tendo como função objetivo a minimização do erro *Root Mean Squared Error*. Também foi estudado qual das bases temporais permitia obter os melhores resultados na previsão do valor das ações, nomeadamente considerando apenas os dados do dia anterior ou os dados dos três dias anteriores à previsão. Para efeitos de comparação e análise dos resultados obtidos foi construído um modelo de persistência para o qual são consideradas para o valor de previsão, o valor da ação no dia anterior. Desta forma, consegue-se concluir quais os modelos que apresentam melhores resultados quando se verifica que o *Root Mean Squared Error* dos modelos é menor que o *Root Mean Squared Error* do seu modelo de persistência.

Por fim tecem-se as conclusões relativamente aos resultados obtidos. De facto, foram encontrados modelos capazes de prever o índice *SP500* ou mesmo ações de empresas presentes no índice *SP500*, mas porque os resultados não se estendem a todas as ações analisadas, essa hipótese não pode ser totalmente corroborada, ainda que dos 8 ativos estudados, 87,5% obteve-se um modelo capaz de prever o preço de ativo melhor que o modelo de persistência. Relativamente às diferenças entre o melhor modelo de cada ativo e o modelo de persistência é possível aferir que as ações com menor volatilidade apresentam diferenças menos significativas.

PALAVRAS-CHAVE

Data mining, machine learning, KNIME, mercado de ações, índice SP500

página propositadamente em branco

ABSTRACT

The present dissertation, carried out in the scope of the master's in industrial engineering and management (MEGI), is the result of an individual project, with the objective of predicting the stock market for a short-term time base, using machine learning models. Only stocks present in the SP500 roughly from 2013 to 2018 were considered for this study.

Although stock market forecasting is a challenge that has constantly motivated the scientific community to search for answers since its existence, it is only in recent years with the existing high computational capacity that there have been some models capable of profitably forecasting the stock market (Rouf et al., 2021). It is noteworthy that even these successful cases still show many weaknesses in terms of consistency (Rouf et al., 2021). The purpose of this project is to analyze a set of macro-economic data to conclude if the present factors allow achieving better forecast results than a persistence model, which considers as forecast value the previous day's closing price.

To carry out this project we followed the CRISP-DM methodology, widely used in data mining projects. Through this methodology it was possible to define clear work objectives, describe and analyze the dataset used, prepare the data considering the models used, submit the dataset to the three models used, namely the Linear Regression, Random Forest Regression and MultiLayer Perceptron, and finally, draw conclusions based on the results obtained.

In the model application phase, it was also taken into consideration to perform sensitivity analyses regarding the parameter values to be used, having as objective function the minimization of the Root Mean Squared Error. It was also studied which of the time bases allowed for the best results in stock price forecasting, namely considering only data from the previous day or data from the three days prior to the forecast. To compare and analyze the results obtained, a persistence model was built for which the forecast value considers the stock value on the previous day. This way, it is possible to conclude which models present the best results when it is verified that the Root Mean Squared Error of the models is smaller than the Root Mean Squared Error of their persistence model.

Finally, conclusions are drawn regarding the results obtained. In fact, models capable of predicting the SP500 index or even stocks of companies present in the SP500 index were found, but because the results do not extend to all the analyzed stocks, this hypothesis cannot be fully corroborated, even though of the 8 assets studied, 87.5% obtained a model capable of predicting the asset price better than the persistence model. Regarding the differences between the best model for each asset and the persistence model, it is possible to conclude that stocks with lower volatility present less significant differences.

KEYWORDS

data mining, machine learning, KNIME, stock market, SP500 index

página propositadamente em branco

ÍNDICE

ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABELAS	XIII
LISTAS DE SIGLAS, ABREVIATURAS E SÍMBOLOS	XVII
1. INTRODUÇÃO	1
1.1. Enquadramento e pertinência	1
1.2. Questão e objetivos de investigação.....	2
1.3. Opções metodológicas	3
1.4. Estrutura do documento	3
2. REVISÃO BIBLIOGRÁFICA.....	5
2.1. Mercado de Ações.....	5
2.1.1. Análise Fundamental.....	7
2.1.2. Análise Técnica.....	7
2.2. Análise Macroeconómica	8
2.3. Data Mining	9
2.3.1. Metodologia CRISP-DM.....	9
2.3.2. Modelos de Previsão (Machine Learning).....	12
2.4. Estado da Arte	14
2.4.1. Tipos de Bases de Dados.....	15
2.4.2. Métodos de Machine Learning	15
2.4.3. Desafios e Futuras Considerações	16
2.5. KNIME.....	17
3. MÉTODOS E APLICAÇÃO	19
3.1. Procedimento Aplicado	19
4. RESULTADOS E DISCUSSÃO	21
4.1. Business Understanding	21
4.1.1. Determinação dos Objetivos de Negócio.....	21
4.1.2. Objetivos do Data Mining	23
4.2. Data Understanding	24
4.2.1. Descrição dos Atributos	24
4.2.2. Análise Setorial do SP500.....	29
4.2.3. Correlações Entre Atributos.....	50
4.2.4. Conclusão do Data Understanding	74
4.3. Data Preparation	75
4.4. Modeling	77
4.4.1. Construção dos Modelos	77
4.5. Avaliação dos Modelos - Evaluation.....	86
4.5.1. SP500.....	86

4.5.2. Facebook	87
4.5.3. Ford Motor Company	87
4.5.4. Coca-Cola Company	88
4.5.5. Kinder Morgan	88
4.5.6. Apple	89
4.5.7. Freeport-McMoRan	89
4.5.8. Host Hotels & Resorts	90
4.6. Discussão de resultados	90
5. CONCLUSÃO	93
5.1. Conclusões finais	93
5.2. Limitações e investigação futura	94
REFERÊNCIAS BIBLIOGRÁFICAS	96
APÊNDICE A – Boxplots dos Atributos do Dataset Inicial	103
APÊNDICE B – Gráficos dos Resultados SP500 (Real VS Previsão e Diferença entre Ambos)	112
APÊNDICE C – Gráficos dos Resultados Facebook (Real VS Previsão e Diferença entre Ambos) ..	114
APÊNDICE D – Gráficos dos Resultados Ford Motor Company (Real VS Previsão e Diferença entre Ambos)	116
APÊNDICE E – Gráficos dos Resultados Coca-Cola Company (Real VS Previsão e Diferença entre Ambos)	118
APÊNDICE F – Gráficos dos Resultados Kinder Morgan (Real VS Previsão e Diferença entre Ambos)	120
APÊNDICE G – Gráficos dos Resultados Apple (Real VS Previsão e Diferença entre Ambos)	122
APÊNDICE H – Gráficos dos Resultados Freeport-McMoRan (Real VS Previsão e Diferença entre Ambos)	124
APÊNDICE I – Gráficos dos Resultados Host Hotels & Resorts (Real VS Previsão e Diferença entre Ambos)	126
APÊNDICE J – KNIME (Integração das Bases de Dados)	128
APÊNDICE K – KNIME (Data Understanding)	129
APÊNDICE L – KNIME (Data Preparation)	130
APÊNDICE M – KNIME (Modelin	131

página propositadamente em branco

ÍNDICE DE FIGURAS

Figura 1 - Cotação do DJIA de 1921 até 2021 (Fonte: (MacroTrends, 2021a))	5
Figura 2 - Cotação do SP500 de 1931 até 2021 (Fonte: (MacroTrends, 2021c))	6
Figura 3 - Cotação da NASDAQ desde 1976 até 2021 (Fonte: (MacroTrends, 2021b))	6
Figura 4 - Fases e Tarefas da metodologia CRISP-DM (Fonte: (Ferreira, 2021a))	10
Figura 5 - <i>Interface</i> da plataforma KNIME (Fonte: (Knime, 2019))	18
Figura 6 - Planejamento do Projeto	20
Figura 7 - Performance e Volume do setor <i>Communication Services</i>	30
Figura 8 - Performance da ação da empresa <i>Facebook</i>	31
Figura 9 - Performance e Volume do setor <i>Consumer Discretionary</i>	31
Figura 10 - Performance da empresa <i>Ford Motor Company</i>	33
Figura 11 - Performance e Volume do setor <i>Consumer Staples</i>	33
Figura 12 - Performance da empresa <i>Coca-Cola Company</i>	35
Figura 13 - Performance e Volume do setor <i>Energy</i>	35
Figura 14 - Performance da empresa <i>Kinder Morgan</i>	36
Figura 15 - Performance e Volume do setor <i>Financials</i>	37
Figura 16 - Performance da empresa <i>Bank of America</i>	38
Figura 17 - Performance e Volume do setor <i>Health Care</i>	39
Figura 18 - Performance da empresa <i>Pfizer</i>	41
Figura 19 - Performance e Volume do setor <i>Industrials</i>	41
Figura 20 - Performance da empresa <i>General Electric</i>	43
Figura 21 - Performance e Volume do setor <i>Information Technology</i>	43
Figura 22 - Performance da empresa <i>Apple</i>	45
Figura 23 - Performance e Volume do setor <i>Materials</i>	46
Figura 24 - Performance da empresa <i>Freeport-McMoRan</i>	47
Figura 25 - Performance e Volume do setor <i>Real Estate</i>	47
Figura 26 - Performance da empresa <i>Host Hotels & Resorts</i>	48
Figura 27 - Performance e Volume do setor <i>Utilities</i>	49
Figura 28 - Performance e Volume da empresa <i>Exelon</i>	50
Figura 29 - Performance do <i>SP500</i> e preços de fecho do <i>SP500</i> (Vermelho), <i>DJIA</i> (Azul) e <i>NASDAQ</i> (Verde) normalizados	51
Figura 30 - Correlações com o índice <i>SP500</i>	51
Figura 31 - Correlações Fortes com o <i>SP500</i> (Azul): <i>GBP/USD</i> (amarelo), <i>ETH</i> (Verde) e <i>FED</i> (Vermelho)	52
Figura 32 - Correlações Muito Fortes com o <i>SP500</i> (Azul Ciano): <i>DJIA</i> (Lilás), <i>NASDAQ</i> (Amarelo), <i>T10Y2Y</i> (Verde), Taxa de Desemprego(Azul) e <i>CSUSHPINSA</i> (Vermelho)	52
Figura 33 - Correlações com o setor <i>Communication Services</i>	53
Figura 34 - Correlações com a empresa <i>Facebook</i>	54
Figura 35 - Correlações com o setor <i>Consumer Discretionary</i>	55
Figura 36 - Correlações com a empresa <i>Ford Motor Company</i>	56
Figura 37 - Correlações com o setor <i>Consumer Staples</i>	57
Figura 38 - Correlações com a empresa <i>Coca-Cola Company</i>	58
Figura 39 - Correlações com o setor <i>Energy</i>	59

Figura 40 - Correlações com a empresa <i>Kinder Morgan</i>	60
Figura 41 - Correlações com o setor <i>Financials</i>	61
Figura 42 - Correlações com a empresa <i>Bank of America</i>	62
Figura 43 - Correlações com o setor <i>Health Care</i>	63
Figura 44 - Correlações com a empresa <i>Pfizer</i>	64
Figura 45 - Correlações com o setor <i>Industrials</i>	65
Figura 46 - Correlações com a empresa <i>General Electric</i>	66
Figura 47 - Correlações com o setor <i>Information Technology</i>	67
Figura 48 - Correlações com a empresa <i>Apple</i>	68
Figura 49 - Correlações com o setor <i>Materials</i>	69
Figura 50 - Correlações com a empresa <i>Freeport-McMoRan</i>	70
Figura 51 - Correlações com o setor <i>Real Estate</i>	71
Figura 52 - Correlações com a empresa <i>Host Hotels & Resorts</i>	72
Figura 53 - Correlações com o setor <i>Utilities</i>	73
Figura 54 - Correlações com a empresa <i>Exelon</i>	74

página propositadamente em branco

ÍNDICE DE TABELAS

Tabela 1 - Estatísticas dos Atributos	28
Tabela 2 - Performance e Volume médio das empresas do setor <i>Communication Services</i>	30
Tabela 3a - Performance e Volume médio das empresas do setor <i>Consumer Discretionary</i>	31
Tabela 3b - Performance e Volume médio das empresas do setor <i>Consumer Discretionary</i>	32
Tabela 3c - Performance e Volume médio das empresas do setor <i>Consumer Discretionary</i>	33
Tabela 4 - Performance e Volume das empresas do setor <i>Consumer Staples</i>	34
Tabela 5 - Performance e Volume das empresas do setor <i>Energy</i>	36
Tabela 6a - Performance e Volume das empresas do setor <i>Financials</i>	37
Tabela 6b - Performance e Volume das empresas do setor <i>Financials</i>	38
Tabela 7a - Performance e Volume das empresas do setor <i>Health Care</i>	39
Tabela 7b - Performance e Volume das empresas do setor <i>Health Care</i>	40
Tabela 8a - Performance e Volume das empresas do setor <i>Industrials</i>	41
Tabela 8b - Performance e Volume das empresas do setor <i>Industrials</i>	42
Tabela 8c - Performance e Volume das empresas do setor <i>Industrials</i>	43
Tabela 9a - Performance e Volume das empresas do setor <i>Information Technology</i>	44
Tabela 9b - Performance e Volume das empresas do setor <i>Information Technology</i>	45
Tabela 10 - Performance e Volume das empresas do setor <i>Materials</i>	46
Tabela 11 - Performance e Volume das empresas do setor <i>Real Estate</i>	48
Tabela 12a - Performance e Volume das empresas do setor <i>Utilities</i>	49
Tabela 12b - Performance e Volume das empresas do setor <i>Utilities</i>	50
Tabela 13 - Quadro Resumo de Análise	75
Tabela 14 - Variáveis para o <i>dataset</i> do <i>Data Understanding</i>	76
Tabela 15 - Análise de Sensibilidade do <i>Random Forest</i> no <i>SP500</i>	78
Tabela 16 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> no <i>SP500</i>	78
Tabela 17 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Facebook</i>	79
Tabela 18 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Facebook</i>	79
Tabela 19 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Ford Motor Company</i>	80
Tabela 20 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Ford Motor Company</i>	80
Tabela 21 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Coca-Cola Company</i>	81
Tabela 22 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Coca-Cola Company</i>	81
Tabela 23 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Kinder Morgan</i>	82
Tabela 24 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Kinder Morgan</i>	82
Tabela 25 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Apple</i>	83
Tabela 26 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Apple</i>	83
Tabela 27 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Freeport-McMoRan</i>	84
Tabela 28 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Freeport-McMoRan</i>	84
Tabela 28 - Análise de Sensibilidade do <i>Random Forest</i> na <i>Host Hotels & Resorts</i>	85
Tabela 29 - Análise de Sensibilidade do <i>MultiLayer Perceptron</i> na <i>Host Hotels & Resorts</i>	85
Tabela 30 - Resultados do <i>SP500</i>	87
Tabela 31 - Resultados da <i>Facebook</i>	87
Tabela 32 - Resultados da <i>Ford Motor Company</i>	88
Tabela 33 - Resultados da <i>Coca-Cola Company</i>	88

Tabela 34 - Resultados da <i>Kinder Morgan</i>	89
Tabela 35 - Resultados da <i>Apple</i>	89
Tabela 36 - Resultados da <i>Freeport-McMoRan</i>	90
Tabela 37 - Resultados da <i>Host Hotels & Resorts</i>	90

página propositadamente em branco

LISTAS DE SIGLAS, ABREVIATURAS E SÍMBOLOS

Lista de Siglas e Abreviações

ANN	Artificial Neural Network
AUD	Australian Dollar
BTC	Bitcoin
CAD	Canadian Dollar
CHF	Confoederatio Helvetica Franc (Swiss Franc)
CNY	Chinese Yuan Renminbi
CRISP-DM	Cross Industry Standard Process for Data Mining
CSUSHPINSA	S&P/Case-Shiller U.S. National Home Price Index
DAX	Deutscher Aktien Index
DEM	Departamento de Engenharia Mecânica
DJIA	Dow Jones Industrial Average
D/E	Debt to Equity Ratio
EPS	Earnings Per Share
ETH	Ethereum
EUR	Euro
FED	Federal Reserve
FOREX	Foreign Exchange Market
FRED	Federal Reserve Economic Data
GBP	British Pound Sterling
JPY	Japanese Yen
IPP	Instituto Politécnico do Porto
ISEP	Instituto Superior de Engenharia do Porto
LTC	Litecoin
MACD	Moving Average Convergence Divergence
MAPE	Mean Absolute Percentage Error
MEGI	Mestrado em Engenharia e Gestão Industrial
MLP	Multilayer Perceptron
MSD	Mean Signed Difference
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
NZD	New Zealand Dollar
PBV	Price to Book Value
P/E	Price to Earnings Ratio
P/S	Price to Sales Ratio
REITS	Real Estate Investment Trusts
RMSE	Root Mean Squared Error

RNN	Recurrent Neural Network
ROA	Return On Assets
ROE	Return On Equity
RSI	Relative Strength Index
SP500	Standard and Poor's 500
SVM	Support Vector Machine
TI	Tecnologias de Informação
T10Y2	10-Year Treasury Constant Maturity Minus 2-Year
USD	United States Dollar

Lista de Símbolos

R^2	Coeficiente de Determinação
-------	-----------------------------

página propositadamente em branco

1. INTRODUÇÃO

Neste capítulo pretende-se fazer um enquadramento do trabalho realizado, identificando a sua pertinência face ao panorama atual, salientando quais os objetivos que se pretendem alcançar com a investigação e por fim evidenciando as metodologias adotadas para alcançar o sucesso dos mesmos.

1.1. Enquadramento e pertinência

O mercado de ações é o termo dado à atividade nas bolsas de valores que tem como principal missão criar mais capitalização para as empresas. Quando uma empresa pretende obter financiamento público e cumpre todos os requisitos que lhe são exigidos, esta é listada numa bolsa de valores permitindo que qualquer pessoa obtenha títulos da empresa. Este fluxo de caixa extra permite que a empresa crie uma estratégia mais ambiciosa, podendo assim crescer de uma forma mais rápida e de forma sustentável. Por outro lado, o investidor ganha dinheiro com a possível valorização do preço de cotação e com o pagamento de dividendos, se praticável (B. Mark Smith, 2003).

Em 1792 é assinado o documento "*Buttonwood Agreement*" que marca a fundação da *New York Stock Exchange* (NYSE), a primeira bolsa de valores oficial nos Estados Unidos. É também aqui que começa o início de uma comunidade de investimento americana que viria a ser conhecida como "*Wall Street*" (Segal, 2021). Desde a sua criação, os mercados financeiros tem vindo a crescer exponencialmente e atualmente, como reportado pelo *The World Bank*, a capitalização de mercado mundial de todas as empresas listadas em bolsas reguladas está perto de atingir o recorde de 100 triliões de dólares (World Federation of Exchanges, 2020).

Devido à sua popularidade, muitas são as tentativas quanto à previsão de determinadas ações ou índices de mercado. A previsão destes mercados é uma tarefa complexa, que exige conhecimento especializado e compilação de grandes quantidades de informação. Para além disso apresenta padrões estocásticos não lineares, dinâmicos, variáveis no tempo e facilmente afetados por muitos fatores económicos e políticos. Gupta (2013) enumera alguns desses parâmetros, como o valor de mercado da ação, a performance de uma empresa e do mercado a que a mesma está envolvida, políticas governamentais, o PIB do país, a taxa de inflação e até fenómenos naturais. A hipótese do mercado eficiente e do *Random Walk* demonstram que o mercado segue um caminho aleatório e é significativamente afetado por novas informações ou notícias (E. F. Fama, 1995). Todo este processo tem despertado à comunidade científica e a alguns especialistas um enorme interesse no sentido de analisar fundamentais que permitam uma maior obtenção de lucros face a uma exposição de risco menor. Existem essencialmente três abordagens que permitem prever os futuros movimentos do preço de uma determinada ação, que são a análise fundamental, a análise técnica e a análise macroeconómica, sendo as primeiras duas as mais frequentemente utilizadas (Jeffrey A. Franke; Kenneth A. Froot, 1990), (Tan, Quek, & Ng, 2007). No entanto é relevante enunciar que tanto o conhecimento especializado, como o processamento de grandes quantidades de dados nem sempre é possível de obter.

Com o desenvolvimento das tecnologias de inteligência artificial ou *machine learning* nos últimos anos, a extração e representação do conhecimento e modelação do mercado com recursos

computacionais provou produzir muito bons resultados neste domínio. Existem já sistemas com uma determinada robustez comprovando que os dados históricos do mercado conjuntamente com os dados extraídos de notícias ou até redes sociais, conseguem prever as alterações na economia e nos setores de negócio com alguma robustez (Rouf et al., 2021). Esta computação integra várias disciplinas, desde a matemática, a economia até à computação numérica, o que tem para isso estimulado esforços de investigação multidisciplinares (Northover, 1999). Contudo, ainda assim existe uma grande margem de progressão, especialmente com o aparecimento de técnicas de aprendizagem profunda.

Com esta dissertação, pretende-se prever o mercado de ações através de fatores macroeconómicos, usando modelos de previsão robustos que permitam atingir resultados positivos quando comparados com o seu respetivo modelo de persistência. Para isso será utilizado um dataset com os históricos de preços das cotações das ações, assim como de fatores macroeconómicos que submetidos aos vários algoritmos de machine learning irão tecer as devidas previsões.

Para o desenvolvimento deste projeto será utilizado a ferramenta KNIME, uma ferramenta *low-code*, de aprendizagem e usabilidade simples e que está disponível gratuitamente. Esta ferramenta permitirá estudar a base de dados, construir toda a lógica do projeto e por fim avaliar os resultados obtidos.

Este trabalho poderá revelar-se importante, na medida que disponibilizará uma reflexão quanto às correlações macroeconómicas dos atributos e posterior modelo de previsão que permitirá à sociedade em geral ter uma visão da potencialidade destas ferramentas.

1.2. Questão e objetivos de investigação

A questão se pretende ver respondida com nesta dissertação é: Com que precisão será possível prever quer o valor de cotação de uma determinada empresa inserida no índice *SP500* quer o próprio índice *SP500* para o dia seguinte, usando essencialmente dados macroeconómicos?

Desta forma, pretende-se criar um modelo, baseado em fatores macroeconómicos fortemente correlacionáveis, capaz de prever o mercado de ações em uma plataforma *low-code*, o KNIME *workflow*. Para isso executar-se-á uma análise regressiva, que permitirá saber com uma determinada incerteza o valor de fecho de uma determinada ação inserida no índice *SP500* para o dia seguinte.

Para além disso, nesta investigação espera-se reconhecer a potencialidade de prever o mercado de ações utilizando uma plataforma *low-code*, o que demonstrará a não exigibilidade de um conhecimento aprofundado em programação para alcançar bons resultados neste campo de estudo. Até ao momento a maioria destes estudos tem sido baseadas em linguagens de programação tais como *python*, mais frequentemente utilizada, e esta investigação espera assim dar uma nova visão quanto à capacidade das plataformas *low-code*.

De forma a alcançar estes objetivos gerais espera-se:

- Analisar e caracterizar um conjunto de dados, nomeadamente a relação entre os atributos e o valor de uma determinada ação do índice *SP500* e do próprio índice *SP500*. Os atributos utilizados serão fatores macroeconómicos, tais como mercado cambial, cripto ativos, *commodities* ou mesmo índices que pretendem refletir o poder económico atual, como por exemplo a taxa de inflação.
- Limpar a base de dados, eliminando todo o tipo de variáveis e valores que poderão induzir o modelo de previsão em erro.
- Preparar a base de dados para que esta seja adequada aos modelos de previsão utilizados.
- Avaliar e comparar diferentes modelos e selecionar o melhor utilizando uma métrica

1.3. Opções metodológicas

A metodologia de investigação frequentemente utilizada pela comunidade científica em trabalhos na área é denominada por CRISP-DM, que permite uma melhor compreensão dos dados, providenciando um modelo completo para conduzir qualquer projeto de *data mining* (Ferreira, 2021b). Esta é composta por seis fases, sendo a primeira o *Business Understanding*, que compreende a análise dos dados extraídos de um conjunto de mercado de ações, a descrição do projeto e dos objetivos a analisar. De seguida o *Data Understanding*, que compreende a compreensão das várias variáveis, assim como eventuais correlações entre elas. Posteriormente faz-se o *Data Preparation*, que visa a descrição do processo de tratamento dos dados, o *Modeling*, que incide na construção do modelo que sustentará a previsão e a *Evaluation* que se refere à análise dos vários modelos de previsão utilizados com intuito de realçar em qual deles obtemos melhor precisão de resultados. Por fim o *Deployment*, que compreende a documentação de suporte.

1.4. Estrutura do documento

A presente projeto está dividido e organizado em cinco capítulos. O presente capítulo, foca-se numa pequena introdução ao projeto abordado, apresentando assim, o enquadramento geral do trabalho, os principais objetivos a serem alcançados e qual a metodologia utilizada para o desenvolvimento do mesmo.

No segundo capítulo é realizada uma revisão bibliográfica relativa aos temas abordados durante o decorrer do projeto, nomeadamente o mercado de ações e suas técnicas de análise, a metodologia CRISP-DM, os tipos de modelos de *machine learning*, o estado da arte e por fim uma pequena abordagem ao KNIME, software *low-code* utilizado. Pretende-se assim, transmitir a base científica que suporta as soluções implementadas.

No terceiro capítulo foi realizado uma descrição mais detalhada sobre o método e abordagens utilizadas ao longo do projeto, descrevendo no final o planeamento das várias etapas do projeto, com o auxílio do *Ganttproject*.

No quarto capítulo é mostrar toda a análise e resultados obtidos neste projeto, onde são seguidos os vários passos da metodologia CRISP-DM. Primeiro discutem-se os objetivos de investigação, depois analisam-se os atributos e a sua correlação com o preço das ações e com o índice de mercado *SP500* e por fim realiza-se a construção dos modelos de previsão analisando para cada um deles os resultados obtidos.

Por último no quinto capítulo são apresentadas todas as conclusões relativas ao trabalho realizado.

2. REVISÃO BIBLIOGRÁFICA

Com o objetivo de dar suporte científico ao trabalho a ser realizado, procedeu-se à pesquisa e posterior descrição dos vários temas abordados. Desta forma, neste capítulo são abordados os conceitos relacionados com o mercado de ações e com a análise fundamental, técnica e macroeconómica. De seguida são abordados os conceitos de *data mining* e de *machine learning*, destacando alguns dos algoritmos que compõem estas áreas de estudo. Posteriormente é analisado o estado da arte, destacando alguns artigos que tem vindo a ajudar a comunidade científica a progredir neste campo de estudo, analisando problemas, objetivos, técnicas de previsão utilizadas e resultados. Por fim, abordou-se a ferramenta KNIME, no sentido de demonstrar as suas vantagens e desvantagens face às linguagens de programação frequentemente utilizadas em projetos *data mining* e *machine learning*.

2.1. Mercado de Ações

O mercado de ações é um instrumento financeiro que permite às empresas obter mais capital para as sustentar a suas atividades de valor acrescentado vendendo quotas de ações da empresa e permite aos seus investidores aumentar o seu capital, quer pelo crescimento do valor determinada ação, quer pelos dividendos, caso a empresa possua essa política (Tretina & Schmidt, 2021). Dentro do mercado de ações, existem bolsas de valores, que são os mercados regulamentados onde se executam todas as transações publicas de títulos financeiros. Nos Estados Unidos as principais bolsas de valores são a NYSE e a NASDAQ e na europa são a London Exchange e a Euronex (Infopédia, 2021).

Um outro conceito é o de índices de ações, que acompanham o desempenho de um grupo de ações (Tretina & Schmidt, 2021). Nas *Figuras 1, 2 e 3* são representados os principais índices de ações dos Estados Unidos, que apresentam uma relevância mundial. Na *Figura 1*, observa-se a cotação do *Dow Jones Industrial Average* (DJIA) também conhecido por *Dow 30* nos últimos 100 anos, que é o índice de mercado que segue as 30 maiores empresas listadas na bolsa de Nova Iorque (NYSE) e na NASDAQ (Akhilesh, Scott, & Schimitt, 2021). Para estarem nestes índices, estas empresas tem de ser economicamente estáveis e serem consideradas *blue-chips*, multinacionais com uma marca respeitável que estão ativas por um número considerável de anos, como é o caso da *Coca-Cola*, *Disney* e *McDonald's* (Chen, Scott, & Munichiello, 2020).

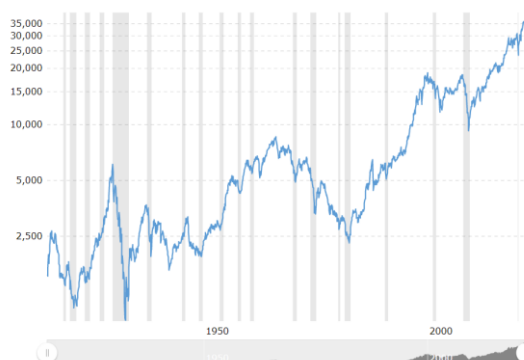


Figura 1 - Cotação do DJIA de 1921 até 2021 (Fonte: (MacroTrends, 2021a))

Na *Figura 2*, observa-se a cotação do *Standard & Poor's 500 Index (SP500)* nos últimos 90 anos, que é o índice de mercado que segue as 500 maiores empresas dos Estados Unidos, sendo apontado como o índice referencia da economia dos estados unidos quando comparado com o *DJIA* (“*Dow Jones Industrial Average*® | S&P Dow Jones Indices,” 2021).

Por outro lado, o cálculo da ponderação difere-se entre os dois índices. O *SP500* utiliza o método de ponderação mais comum, que se baseia na capitalização de mercado, dando assim maior percentagem de atribuição às empresas com maior capitalização de mercado. Já no *DJIA*, utiliza-se um método de ponderação baseado no preço, o que permite alocar maior percentagem nas empresas com preços de ações mais elevados (Kenton, Boyle, & Munichiello, 2021)



Figura 2 - Cotação do SP500 de 1931 até 2021 (Fonte: (MacroTrends, 2021c))

Na *Figura 3*, observa-se a cotação da *NASDAQ* nos últimos 45 anos, que é o índice de mercado que contem todas as empresas que estão inseridas na bolsa *NASDAQ*, sendo a maioria delas empresas do setor tecnológico relacionadas com a internet. Esta, ao contrário da *NYSE*, não tem uma sede presencial, sendo todas as execuções realizadas por computador. Até ao momento estão 3300 empresas inseridas nesta bolsa de valores (The Investopedia team & Anderson, 2021).



Figura 3 - Cotação da NASDAQ desde 1976 até 2021 (Fonte: (MacroTrends, 2021b))

Ambos apresentam uma tendência crescente, o que explica o sucesso dos investimentos a longo prazo nestes mercados. No entanto, é possível constatar nas figuras acima mencionadas grandes quedas que apresentaram uma recuperação longa, registam assim o elevado risco associado ao investimento destes produtos financeiros. A queda acumulada de -86% no *Dow Jones*

relacionada com o colapso da bolsa em 1929 e a queda acumulada de mais de 80% no NASDAQ relacionada com a bolha *dotcom* em 1999 são exemplos do risco envolvido (Rankia Portugal, 2021).

Durante o século 20 desenvolveram-se algumas teorias relativamente à previsão do mercado de ações, sendo a hipótese do mercado eficiente (E. F. Fama, 1970) e teoria do caminho aleatório (Malkiel, 1996) as mais conhecidas. A primeira afirma que o preço de uma determinada ação é baseado na integração de toda a informação disponível, fazendo com que nova informação seja imediatamente descoberta e eliminada por uma correção do preço, o que impede a sua previsão para uma determinada base temporal futura. Nos últimos anos a veracidade desta teoria tem sido refutada, sendo o próprio autor no artigo "*Efficient Capital Market*" uma das pessoas a admitir essa possibilidade (E. F. Fama, 1995). A segunda defende que a previsão do mercado de ações é impossível sendo que os preços são determinados aleatoriamente e o bom desempenho, mesmo com bastante informação à nossa disponibilidade, é inviável e ineficiente (Malkiel, 1996).

Falinouss (2007) menciona algumas abordagens capazes de retornarem resultados positivos nesta área, referindo a abordagem fundamental como aquela que permite identificar as ações subavaliadas face aos fatores financeiros que a mesma apresenta, a abordagem técnica como a melhor para analisar o melhor momento de compra ou venda no curto prazo e a abordagem macroeconómica para uma análise correlativa da ação com o fatores de mercado e/ou económicos. Para esta investigação e consequentemente previsão, usar-se-á exclusivamente a análise macroeconómica, sendo para isso estudadas correlações entre vários ativos ou constantes económicas e o preço de ações ou índice *SP500*.

2.1.1. Análise Fundamental

A análise fundamental baseia-se na evolução dos preços das ações de uma determinada empresa através do seu histórico de dados financeiros e contabilísticos, o que inclui a análise dos ganhos, despesas, lucros, ativos e passivos da empresa, do setor e da economia que esta está sujeita. Desta forma, os analistas conseguem obter bons desempenhos (Mahmoud, A., and Sakr, 2012), (Elleuch, 2009), (E. Fama & Kenneth R. French, 1993), (Piotroski J, 2002). Esta abordagem é utilizada pelos analistas de forma a identificar empresas com potencial crescimento e rendimento no longo prazo e avaliar ações subavaliadas no curto prazo (Muhammad, 2018).

Os fluxos de caixa da empresa afetam a sua liquidez e consequentemente as suas atividades operacionais (Dobson & Soenen, 1993), enquanto o aumento da liquidez ou dos ativos correntes pode também assinalar o desequilíbrio do retorno do investimento face aos objetivos da empresa (Horne & Wachowicz, 2004). Isto pressiona a empresa a fazer ajustamentos aos fluxos de caixa e liquidez da empresa, medidos através dos rácios da análise fundamental, e a apresentar uma excelente estrutura de ativos alinhados com os objetivos financeiros (Muhammad, 2018).

2.1.2. Análise Técnica

A análise técnica envolve o estudo dos preços históricos, utilizando os gráficos como ferramenta principal para detetar essas oscilações. Esta abordagem está mais centralizada nos indicadores de mercado, incluindo análises baseadas nos movimentos dos preços das ações, nas previsões financeiras e no volume e tendência de mercado. O princípio básico desta ferramenta

inclui conceitos como a tendência natural dos preços, confirmações e divergências e o efeito do volume (Hellström & Holmström, 1998).

Por outro lado, a análise técnica ignora qualquer tipo de dados fundamentais financeiros na elaboração de estratégias de investimento (Neely, White, & Street Journal, 1992). Esta é mais atrativa para os investidores de curto prazo, que tem o intuito de determinar o valor atual do mercado e os seus sentimentos sobre as possíveis alterações esperadas num futuro próximo (Allen & Taylor, 1990).

Pring (1991) menciona também que a análise técnica se baseia na análise de uma série de dados numéricos e tenta prever os mercados de ações utilizando indicadores de análise técnica. A hipótese que todas as reações do mercado às notícias estão consideradas nos preços das ações permite ignorar as notícias enquanto fator de previsão. O seu principal foco é assim identificar as tendências históricas antecipando futuras tendências.

Schumaker e Chen (2009) referem que na análise técnica, o tempo de mercado é crítico e as oportunidades podem ser encontradas através de uma cuidadosa averiguação dos preços históricos e do movimento de volume quando comparados com o preço atual. Os analistas utilizam gráficos e modelam técnicas para identificar tendências ou padrões no preço e no volume. Uma das desvantagens desta técnica é que os gráficos apenas contem os eventos e não as causas (Kroha & Baeza-Yates, 2004).

2.2. Análise Macroeconómica

Existem dois tipos de riscos associados aos investimentos, nomeadamente o risco sistemático e o risco não-sistemático, em que ambos podem afetar o lucro esperado pelos investidores. O risco não sistemático de uma empresa não está correlacionado com outras empresas. Inversamente, o risco sistemático, também conhecido como o risco de mercado, correlaciona-se com todas as empresas, afetando assim todo o mercado. Alguns dos factores que influenciam o risco sistemático, também denominados como variáveis macroeconómicas, são por exemplo a inflação, taxas de câmbio, taxas de juro, etc... Consequentemente, as mudanças nas variáveis macroeconómicas terão um impacto sobre as ações de uma determinada empresa (Rizky Sabilla & Kurniasih, n.d.).

A inflação é o aumento continuado dos preços em geral. A inflação ideal é uma inflação moderada que está de acordo com as capacidades económicas do país (Putong. 2013). Hasyim (2016) menciona os efeitos adversos da inflação nos indivíduos e na sociedade, que se traduzem na redução do rendimento das pessoas com rendimento fixo, a redução do valor da riqueza sob a forma de dinheiro e uma maior disparidade entre as várias classes sociais existentes. Os resultados da pesquisa conduzida por Adesanmi e Jatmiko (2017) realçam que a inflação tem um efeito negativo sobre os preços das ações. Em contraste, a investigação de Mulyani (2012) concluiu que a inflação tem um efeito positivo sobre os preços das ações. Mas Rachmawati e Laila (2015) descobriram que a inflação não tem qualquer efeito sobre os preços das ações.

A taxa de câmbio nominal é o valor referencia que se utiliza quando há uma troca entre duas moedas diferentes. O câmbio estrangeiro (FOREX) é um instrumento de pagamento utilizado para financiar transacções económicas internacionais e tem um registo oficial da taxa de câmbio num banco central (Putong. 2013). Os resultados do estudo de Kusuma Badjra (2016) referencia que as taxas de câmbio têm uma influência positiva sobre o índice de preços das ações. Em contraste,

Husnul et. al (2017) descobriram que as taxas de câmbio afectam negativamente os preços das acções.

Os juros bancários de acordo com Kasmir (2016) são definidos como uma remuneração por serviços prestados aos clientes com base em princípios convencionais. As actividades bancárias convencionais têm diariamente dois tipos de juros dados aos seus clientes Kasmir (2016), nomeadamente o depósito juros e juros de empréstimo. Os resultados de Febrina et. al (2018) concluem que as taxas de juro não têm influência sobre os preços das acções. No entanto, pesquisas de Husain (2013) descobriram que as taxas de juro têm uma influência negativa sobre os índices de preços das acções. Em contraste, a pesquisa Rahman P et. al (2017) descobriu que as taxas de juro têm um efeito positivo sobre o preço das acções.

Seng e Hancock (2012) confirma em seu estudo realizado em um período de dez anos para 33 países do mundo, um efeito significativo das variáveis macroeconómicas sobre as empresas, o setor e país na previsão desses retornos.

2.3. Data Mining

Data mining, também denominada por análise do conhecimento de base de dados, extração de conhecimento, análise de dados/padrões e recolha de dados (Han, Kamber, & Pei, 2011), é definido como a extração de informação atípica, desconhecida, e potencialmente útil de uma determinada base de dados (Frawley, Piatetsky-Shapiro, & Matheus, 1992). Ferreira (2021b) define *data mining* como o uso de técnicas eficientes que permitem analisar a existência de padrões a partir de uma grande quantidade de dados. O objetivo de qualquer processo de *data mining* é construir um modelo preditivo ou descritivo de uma grande quantidade de dados que além de explicar certos comportamentos permite a geração de mais informação (Falinouss, 2007).

Data mining surge em 1980 através da interseção de alguns campos de estudo, tais como *machine learning*, reconhecimento de padrões, bases de dados, estatística, inteligência artificial, visualização de dados e computação de elevado rendimento (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Segundo Moro, Laureano, Cortez, e Algoritmi (n.d.) esta disciplina apresenta modelos de previsão nos seguintes campos de estudo:

- Setor financeiro, devido à qualidade e fiabilidade de dados que estes apresentam é possível efetuar previsões sobre pagamentos de empréstimos e até identificar fraudes de cartões de créditos.
- Indústria retalhista, pela necessidade de compreender grandes quantidades de dados sobre vendas, stocks, compras, recomendações e até campanhas promocionais.
- O setor energético, nomeadamente na previsão do preço da eletricidade ou da potência consumida
- Fins meteorológicos, como a previsão do tempo para uma determinada base temporal.

2.3.1. Metodologia CRISP-DM

Um típico processo de *data mining* compreende a utilização da metodologia CRISP-DM, que procura assim ajudar qualquer pessoa a conduzir um projeto de *data mining*. Esta metodologia

consiste em seis fases: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* e *Deployment* (Ferreira, 2021a).

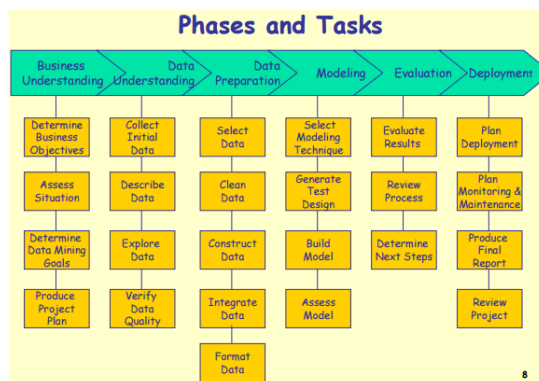


Figura 4 - Fases e Tarefas da metodologia CRISP-DM (Fonte: (Ferreira, 2021a))

Business Understanding é a primeira e talvez a mais importante fase de qualquer projeto de exploração de dados. Nesta fase o foco do analista centra-se em compreender os objetivos do projeto do ponto de vista do negócio, formulando assim os objetivos com a exploração da base de dados e um plano com vista o alcance desses objetivos (Clinton, 2000).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Definição dos objetivos de negócio. O analista procura identificar e compreender quais os objetivos e requisitos do projeto e descobrir fatores importantes que podem influenciar o resultado do projeto. Desta forma, negligenciar este passo é colocar toda a investigação em risco de se tornar obsoleta e improdutiva, pois obteremos as respostas corretas às questões erradas.
- Avaliação dos recursos, restrições, pressupostos, riscos ou outro tipo de fator relevante.
- Definição dos objetivos com a exploração de dados. Estes objetivos apresentam uma maior componente técnica que os objetivos de negócio e por isso são realizados numa fase posterior.
- Desenvolvimento de um plano para os alcançar. O plano deverá especificar os passos a serem realizados, incluindo a seleção dos recursos e metodologias adotadas.

Data Understanding é a fase que permite recolher e compreender os dados. O objetivo é identificar os problemas relacionados com a qualidade dos dados, retirar informação dos dados compreendidos ou explorar subconjuntos de dados com a finalidade de formular hipóteses sobre algum tipo de dados mais específico (Clinton, 2000).

Nesta fase é realizado uma certa sumarização dos dados, que providencia as principais propriedades das variáveis ou também denominadas por atributos. Desta forma, são realizadas análises gráficas através de ferramentas tais como *boxplots* ou histogramas para avaliar medidas como a média, a moda, a mediana, a variância, o desvio padrão, o coeficiente de variação e a simetria. Para análises entre duas variáveis é comum o uso de correlações e covariâncias (Wirth & Hipp, n.d.).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Recolha dos dados iniciais. Os dados obtidos devem ser assegurados através de uma fonte fidedigna, podendo, se assim o entender, efetuar algum tipo de *Data Preparation*.

- Descrição e exploração dos dados. Relativamente à descrição esta permitirá examinar algumas propriedades como o tipo de variável e o intervalo de valores. Já quanto à exploração, esta aborda os objetivos identificados utilizando técnicas de consulta e visualização que permitem refinar algum tipo de dado anteriormente descrito.
- Verificação das qualidades dos mesmos.

Data Preparation inclui todas as atividades que permitirão obter uma base de dados limpa e normalizada, contendo informação útil para alimentar o modelo de previsão sem qualquer tipo de enviesamento. Desta forma, esta é a fase que demora 90% do tempo disponibilizado para o projeto (Clinton, 2000).

Nesta fase é frequentemente utilizado uma técnica denominado por amostragem. Esta técnica permitirá criar um subconjunto orientado aos nossos objetivos, que pela sua propriedade será representativo do conjunto original. Para criar esta amostra, são realizadas ações como o tratamento de valores em falta, *outliers*, valores com pouca ou nenhuma variabilidade e variáveis que tem uma correlação perfeita com a variável objetivo (Wirth & Hipp, n.d.).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Seleção das variáveis a serem utilizadas pelo modelo.
- Limpar os dados, aumentando a qualidade dos dados, tal como pretendido.
- Eventual construção de novas variáveis através das existentes. Este passo consiste na preparação da data considerando o objetivo final que se pretende obter. Podem assim ser realizadas conversões de variáveis nominais para numéricas e vice versa e normalizações de variáveis.

Na fase de *Modeling* o objetivo principal é selecionar e aplicar as várias técnicas de modelação. Como cada modelo tem os seus parâmetros ou requisitos quanto ao tipo de variável pode eventualmente ser necessário voltar ao *Data Preparation* (Clinton, 2000).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Seleção do modelo de previsão. É importante considerar o output do resultado que pretendemos, pois poderá ser regressivo (variável ordinal) ou categórico (variável do tipo nominal).
- Geração de modelo de teste para testar a qualidade e validade do modelo. Tipicamente é executado um modelo de treino que será testado no modelo de teste. A qualidade do modelo é assim medido pela precisão do modelo de teste.
- Construção do modelo e ajustar todos os parâmetros de forma a obter maior eficiência dos mesmos, tendo em conta a função objetivo pretendida.
- Avaliar o modelo em termos técnicos. O uso da função de custo, que consiste na diferença entre a previsão e o real, é uma boa métrica para avaliar o modelo.

Na fase de *Evaluation* é feita uma avaliação ao modelo com a finalidade de avaliar se os objetivos e requisitos de negócio foram alcançados (Clinton, 2000).

Com o intuito de tornar este processo mais eficiente são requeridos o uso de algumas métricas, tais como a matriz de confusão, a matriz de custo, a precisão, a tendência, a variância e a curva ROC (Moro et al., n.d.).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Avaliar os resultados face ao alcance dos objetivos propostos e se caso não o forem, é importante encontrar o motivo. Nesta fase novos desafios ou informação podem surgir e devem ser enunciados para futuras análises.
- Rever o processo com o objetivo de encontrar algum tipo de variável que fora negligenciado.
- Determinar os próximos passos. Avaliar o sucesso do projeto e se o mesmo for positivo poderá se avançar para a próxima e última fase.

A fase do *Deployment* tem como finalidade organizar, documentar e apresentar os resultados do projeto (Clinton, 2000).

Segundo Colin (2000) esta fase do CRISP-DM compreende a:

- Planeamento do *Deployment*. A equipa faz uma avaliação dos resultados e conclui a estratégia de *Deployment*.
- Planeamento de monitorização e manutenção de forma a evitar desvios de informação que afetaram a precisão do modelo
- Execução do relatório final contendo um resumo e apresentação do projeto.
- Retrospectiva do projeto, percebendo o que correu bem, o que correu mal e oportunidades de melhoria.

Colin (2000) refere alguns exemplos na conclusão como o caso da empresa OHRA, onde um novo funcionário utilizou o CRISP-DM para planear e orientar um projeto de *data mining* e foi extramente bem sucedido e o caso da *Chrysler* que adaptou o CRISP-DM para o desenvolvimento da sua própria ferramenta especializada na gestão da relação com o cliente, melhorando assim o seu marketing.

Wirth e Hipp (n.d.) concluem que o CRISP-DM funciona, embora menciona que o nível de detalhe ideal ainda é difícil de reconhecer e o processo de avaliação é por vezes complexo, especialmente em termos de velocidade e custos.

Moro, Laureano, Cortez, e Algoritmi (n.d.) refere a utilização desta metodologia para as campanhas de marketing no setor bancário, que se provou ser bastante útil pois permitiu que o modelo de previsão obtivesse melhores resultados.

2.3.2. Modelos de Previsão (Machine Learning)

Tal como anteriormente referido, é no *modeling* que se seleciona e aplica-se determinados modelos. Em *machine learning* existem dois tipos de modelos. Os modelos supervisionados, classificados pelo tipo de dados como sendo de classificação e regressão, definem os dados de input e os resultados são dados através de algoritmos de aprendizagem. Já os modelos não supervisionados, classificados como de agrupamento, associação e sumarização, não definem os dados de entrada, sendo o algoritmo de aprendizagem responsável por definir o padrão que permitirá obter o output desejado. Neste trabalho serão utilizados apenas modelos supervisionados de regressão.

Dentro destes, existem os métodos *Distance-based*, que são essencialmente o *Nearest Neighbour*, o *K-Nearest Neighbour* e o *Case-Based Reasoning*. O *Nearest Neighbour* reconhece e armazena o output dos valores históricos para prever o output do valor desconhecido, através da

distância entre eles, podendo ser aplicado quer em modelos de classificação, quer em regressão. Neste caso, a normalização dos dados de treino favorece a precisão do modelo. Já o *K-Nearest Neighbour* tem o mesmo princípio do anterior, mas com um parâmetro k que deve ser um número ímpar e o menor possível. As suas vantagens são o facto de memorizar os objetos de treino, simplificar o output de objetos complexos com aproximações e obter uma natureza incremental. Por outro lado, tem como desvantagens, a ineficiência dependendo da quantidade de dados e a sensibilidade a atributos irrelevantes. Por último o *Case-Based Reasoning* resolve um novo problema baseado na solução de problemas similares. Desta forma, dado um problema, ele memoriza-o, procura por problemas similares, mapeia a solução do novo problema baseado na do previamente registado e por fim armazena a solução do problema na sua memória (Ferreira, 2021c).

Relativamente aos métodos probabilísticos, os mais conhecidos são o *Bayesian Learning*, o *Naive Bayes* e o *Bayesian Networks*. O *Bayesian Learning* determina a probabilidade de um evento A sabendo B , onde A é a classe e B as observações teste dos atributos (Ferreira, 2021c). Já o *Naive Bayes* considera que os atributos de uma amostra são independentes da sua classe. Apresenta assim um bom desempenho mesmo que os atributos não sejam independentes e é robusto perante atributos irrelevantes e com ruído e os resultados (Xianya, Mo, & Haifeng, 2019). Enquanto o *Bayesian Networks* utiliza a independência condicional para encontrar um equilíbrio entre o número de parâmetros calculados e a representação das dependências entre as variáveis (Ferreira, 2021c).

Quanto aos métodos de pesquisa, é relevante abordar a *Decision Tree* e a *Regression Tree*. Neste método de estrutura hierárquica, um exemplo é submetido a uma série de testes que determinarão a sua previsão. Os exemplos de treino são usados de forma a obter os testes mais apropriados e os testes que mais maximizem a precisão da previsão são selecionados primeiro. A *Decision Tree* procura uma medida, denominada por entropia, que prefira atributos com um grande grau de ordem, ou seja, que a maioria dos exemplos seja da mesma classe. Já a *Regression Tree*, uma variante da *Decision Tree* tem como objetivo aproximar funções com valor real. Este método é construído através do processo iterativo de partição recursiva binária, que vai dividindo os subconjuntos em subconjuntos cada vez mais pequenos (Whalen, Prakash Pandey, & Pandey, n.d.).

O modelo *Random Forest Regression* é uma extensão do modelo *Decision Tree Regression* criado para colmatar a maior desvantagem do último, que é o *over-fitting*. O *Random Forest* usa a média do resultado de múltiplas *Decision Trees*, como uma espécie de votação, para gerar o resultados final de forma aleatória. É também conhecido como um modelo Ensemble, métodos que utilizam as melhores características de modelos mais básicos para obtenção de melhores resultados, formando assim modelos mais robustos e complexos. Em termos de regressão, o algoritmo segue o conceito da regressão linear (Breiman, 1994).

Relativamente aos métodos de otimização, é relevante abordar as *Neural Networks* e o *Support Vector Machine*. O *Artificial Neuron (ANN)* é um método baseado nos neurónios biológicos, ou seja, cada neurónio recebe um determinado input que depois de processado gera um determinado output. As vantagens das *Artificial Neural Networks* são a precisão da classificação, que mesmo para problemas complexos é alta, a robustez em lidar com informação irrelevante ou imprecisa e os resultados poderem do tipo discreto ou reais (Sharma et al., 2021). Já as desvantagens das *Artificial Neural Networks* são a dificuldade em determinar uma topologia de rede ótima para um

problema, a possibilidade de demorar muito tempo a encontrar uma solução adequada, a complexidade devido aos parâmetros que possui e o facto de não providenciar explicações dos resultados (Ganser, Hollaus, & Stabinger, 2021). Já as *Recurrent Neural Networks* (RNN) permitem lidar com dados sequenciais. Estas tem como desvantagens a inabilidade em reter informação quando a sequencia é longa e o facto de esquecer informação mais antiga na sua memória, o que limita o seu desempenho (Ferreira, 2021c). Por fim o *Support Vector Machine* (SVM) é um novo método de classificação que admite dados lineares e não lineares. Através de um mapeamento não linear transforma os dados de treino originais numa dimensão superior, onde procura pela fronteira de decisão, limitando o erro e aumentando as margens geométricas (Farhat, 1992).

O modelo *MultiLayer Perceptron* é uma ANN que possui três ou mais camadas, sendo assim considerado uma ANN de aprendizagem profunda. Estas camadas dividem-se em uma única de entrada, uma ou mais escondidas e uma única de saída, sendo que a fluxo é unidirecional, ou seja, do input para o output. Os neuroes são treinados através de um algoritmo de retro propagação. Tem como vantagens o facto de resolverem problemas estocásticos, que permite obter soluções viáveis para problemas complexos (Murtagh, 1991).

Relativamente aos *Combining Multiple Learners*, estes tem como objetivo combinar modelos bases que fazem diferentes tipos de erros, de forma a minimizá-los. O *Bagging – Bootstrap Aggregating* cria subconjuntos dos dados de treino de forma aleatória, treina os dados nesses subconjuntos e por fim efetua a previsão do conjunto fazendo uma agregação do modelo (Breiman, 1994). Desta forma, diminui o erro diminuindo a variância nos resultados. Já o *Random Forest* é uma variação do algoritmo *Bagging* que usa um conjunto de *Decision Trees*, sendo o número de conjuntos um parâmetro a determinar (Ferreira, 2021c).

O *Boosting* é um processo iterativo para mudar a distribuição dos dados de treino, que se foca nos registos classificados erradamente. O *Boosting* fornece assim uma maior precisão do que o *Bagging*, embora seja menos consistente. Por outro lado, o *Boosting* está particularmente sujeito a *overfitting*, ou seja, baseia-se exclusivamente num determinado subconjunto de treino para determinar a previsão (Ferreira, 2021c).

O *Staking* envolve o treino de um algoritmo de aprendizagem para combinar as previsões prévias de outros algoritmos. Todos os outros algoritmos são treinados primeiro usando os dados disponíveis, depois são agregados a um algoritmo que faz a previsão final utilizando todas as previsões anteriores como inputs. Contrariamente ao *Boosting* e ao *Bagging*, este algoritmo é mais usado para gerar conjuntos de classificadores heterogéneos, ou seja, conjuntos que usam diferentes algoritmos de indução (Ferreira, 2021c).

O modelo Linear Regression é um dos modelos mais conhecidos em estatística e machine learning, que tem bastante popularidade por ser simples de representar. O algoritmo de regressão linear mostra uma relação linear entre uma variável dependente (y) e uma ou mais variáveis independentes (x). Desta forma, a regressão linear mostra a relação linear, ou seja, descobre como o valor da variável dependente está a mudar de acordo com o valor da variável independente (Su, Yan, & Tsai, 2012).

2.4. Estado da Arte

De forma a avaliar o estado da arte serão mencionados alguns estudos na área, abordando as respetivas bases de dados e variáveis utilizadas, assim como avaliação dos resultados obtidos.

Verifica-se o desenvolvimento de estudos nesta área quer para ações, como é o caso da *Apple* (Weng, Ahmed, & Megahed, 2017), Google (Di Persio & Honchar, 2017), quer em índices, tal como o *DJIA* (Ranco, Aleksovski, Caldarelli, Grčar, & Mozetič, 2015), o *SP500* (Yudong & Lenan, 2009), *NASDAQ* (Guresen, Kayakutlu, & Daim, 2011) e o *DAX* (Lugmayr & Gossen, 2012).

2.4.1. Tipos de Bases de Dados

Para os vários estudos que envolvem a previsão do mercado de ações, são utilizados bases de dados que contemplam os dados históricos relacionados com análise fundamental, análise técnica e análise macroeconómica, pois tal como evidenciado anteriormente, eles refletem a informação necessária para a compreensão do comportamento do mercado. Embora a maioria dos estudos encontrados neste campo de investigação contempla a análise técnica ou a análise fundamental (Jeffrey A. Franke; Kenneth A. Froot, 1990), (Tan et al., 2007).

Principalmente para estudos com uma base temporal de curto prazo, os indicadores técnicos apresentam grande relevância para o sucesso da previsão, devido à sua representação sintetizada das tendências em séries de dados temporais. Alguns desses exemplos são os indicadores de tendência, os indicadores de momento, os indicadores de volatilidade e os indicadores de volume (Devi & Bhaskaran, 2015).

A maioria dos dados de mercado são gratuitos e podem ser descarregados diretamente de websites dedicados à disponibilização de informação financeira, como é o caso da *Yahoo Finance* (Weng et al., 2017), (Dey, Kumar, Saha, & Basak, 2016), (Nayak, Pai, & Pai, 2016) e *Bloomberg* (Milosevic, 2016).

Quanto à base temporal tem se verificado previsões para todo o tipo de investidores, sejam de curto ou longo prazo. Todos os estudos envolvem uma base temporal fixa que pode ser inferior a um dia (Huang & Li, 2017), diária (Peachavanish, 2016), semanalmente (Shah, Isah, & Zulkernine, 2019), mensal (Nayak et al., 2016) e por aí em diante.

A maioria dos estudos são baseados em previsões categóricas, permitindo assim reconhecer se uma determinada ação para uma determinada base temporal tem um comportamento crescente, decrescente ou inalterado (Devi & Bhaskaran, 2015).

2.4.2. Métodos de Machine Learning

Nesta secção pretende-se abordar os modelos de *machine learning* mais utilizados nesta área de estudo.

Guresen (2011) menciona o uso de modelos ANN para prever o índice *NASDAQ* para um horizonte temporal de dois em dois dias. Zhong e Enke (2017) também referem o uso destes modelos para a previsão diária do *SP500*. Já Weng (2017), Di Persio e Honchar (2017) mencionam o uso de MLP, um tipo de método derivativo das ANN. Geralmente o MLP os erros da previsão provenientes do output são novamente inseridos como inputs de forma a minimizar os erros (Goh, 1995), (Zhang et al., 2021). Milosevic (2016) fez um estudo de previsão aos 5 maiores índices do mundo, o *DAX*, *DJIA*, *FTSE-100*, *HSI* e *NASDAQ*. Eles concluíram que as ANN foram capazes de prever o valor da cotação com uma precisão superior a 60% (Milosevic, 2016).

O método mais popular para estudos da área é, tal como referido anteriormente, o SVM (Jagla, Wiswedel, & Coppée, 2011), (Nti, Adekoya, & Weyori, 2020), (Dey et al., 2016), (Milosevic, 2016). Embora (Nayak et al., 2016) desenvolveu um estudo que envolve duas bases temporais diferentes, nomeadamente diário e mensal. Para esse estudo utilizou dados provenientes dos setores bancários e petrolíferos que alimentaram três algoritmos, nomeadamente a *Decision-Boosted Tree*, o SVM e a *Logistic Regression*, sendo utilizada a precisão como principal métrica de avaliação da performance. Aquilo que se comprovou é que a *Decision-Boosted Tree* superaram os restantes métodos, atingindo precisões de 54,8% para dados provenientes do setor bancário e 76,9% para dados provenientes do setor petrolífero. Já o SVM para os respetivos setores atingiu precisões de 51% e 44,2%.

Milosevic (2016) refere o método *Random Forest* como o método que obteve melhor performance no seu estudo, que envolveu a precisão de várias ações dos principais índices mundiais. Este método obteve uma precisão de 75,1%, enquanto a *Decision Tree* obteve 68,7%, o SVM 63,9% e o *Naive Bayes* 54,5%. Este último, apesar de neste estudo obter menos precisão é referenciado pelo autor como sendo bastante popular neste tipo de estudos devido a sua execução rápida no processamento de grandes quantidades de dados (Milosevic, 2016), (More, Rathod, Patil, Sarode, & Student, 2018), (Xianya et al., 2019).

Por fim, quanto aos modelos de regressão, que contrariamente aos modelos categóricos conseguem evidenciar a previsão numérica de uma determinada ação, podem-se destacar o *Simple Linear Regression* (Gururaj & R, 2019), o *Multiple Regression* (Enke, Grauer, & Mehdiyev, 2011), a *Decision Tree Regression* (Nti et al., 2020), a *Logistic Regression*, o *Support Vector Regression* (Shah et al., 2019) e o *Ensemble Regression* (Nti et al., 2020). Ananthi e Vijayakumar (2020) desenvolveram um modelo que prevê o preço de uma empresa inserida pelo utilizador. Através de uma combinação dos modelos anteriores foi possível chegar a uma precisão de 85%.

Em termos de avaliação, a maioria dos estudos provenientes destas áreas utiliza a *Accuracy* como métrica de avaliação dos modelos de previsão, que para os estudos referenciados varia entre 50 e 90%. A *Accuracy* é dada pela percentagem de previsões corretas sobre o número total de instâncias de teste (Rouf et al., 2021).

2.4.3. Desafios e Futuras Considerações

Atualmente cada vez mais o acesso à informação é facilitado, mas por outro lado, a exatidão dos dados nem sempre é a melhor e por vezes verifica-se uma dificuldade em extrair a informação realmente necessária para uma melhor precisão do modelo a contruir (Robles, Ivan V, & Belik, 2010).

Alguns dos autores pronunciam se quanto à previsão das ações em tempo real utilizando as metodologias anteriormente referidas. No entanto, é mencionado que estes métodos podem apenas funcionar em circunstâncias controladas, pois surgem fatores como a variação do preço, ruído e eventos imprevistos, não deixando de ser um desafio para eventuais estudos futuros (Rouf et al., 2021).

O mercado de ações é um mercado muito volátil, ou seja, apresenta uma enorme flutuação de preço para bases temporais curtas. Algumas das razões para essa volatilidade são a incerteza e a inflação, que fazem aumentar o risco de investimento. Esta característica faz com que a previsão

do mercado de ações seja desafiante, muitas das vezes estando intimamente correlacionado com o sucesso da previsão, ou seja, quanto mais volatilidade, menor é a precisão da previsão (Rouf et al., 2021).

Cada vez mais se assistem a eventos de pânico no mercado, devido quer à volatilidade referenciada anteriormente, quer à especulação existente no mercado. Esta caracteriza-se por uma reação exagerada do mercado, que é no fundo uma característica psicológica do ser humano face ao medo e à perda, faz com o processo de previsão do mercado de ações seja ainda mais desafiante (Shah et al., 2019).

A parte fascinante desta área de investigação é a sua natureza autodestrutiva. Isto significa que partilhar as metodologias que geram os lucros mais elevados e por consequência tem maiores precisões face aos demais estudos tornarão no futuro essas metodologias inúteis. Desta forma, os melhores algoritmos nesta área de estudo são restritos e privados, nunca sendo publicados. Por outro lado, cada vez mais novos algoritmos estão a ser inseridos no mercado procurando resolver pequenos problemas que os atuais possuem ou não conseguem dar resposta (Rouf et al., 2021).

Através de uma breve pesquisa constata-se que existem poucos estudos que contemplam a previsão do mercado de ações através da análise macroeconómica, o que também se torna um desafio para o futuro desta área científica. Por outro lado, embora cada vez mais se utiliza software low code para a realização de estudos científicos, este ainda é um desafio para algumas áreas, como é o caso da previsão de ativos financeiros. A evolução destes softwares e consequentemente estudos científicos provará que não será preciso um vasto conhecimento em programação para a realização de estudos significativos nesta área, sendo atingidos graus de incerteza reduzido.

2.5. KNIME

A plataforma de análise KNIME é um software intuitivo disponível para toda a comunidade permitindo a criação de aplicações para a ciência dos dados e serviços. Esta ferramenta permite assim compreender uma base de dados através de programação *low-code*, que compreende o uso de nodos com uma logica associada. Estes nodos utilizam características gerais do software, permitindo por exemplo, o tratamento de milhões de linhas num ambiente de trabalho com uma capacidade computacional de apenas 4GB de *Random Access Memory (RAM)*. Para além disso tem vindo a integrar novos desenvolvimentos, permitindo construir fluxos mais robustos (KNIME, 2021c).

Assim como outras plataformas de *low-code* como é o caso da *Outsystems*, esta disponibiliza uma certa quantidade de material didático como livros, cursos, documentação técnica, tutoriais e certificações (KNIME, 2021b). Outra vantagem deve se à comunidade presente no fórum e no separador KNIME *Hub* onde possível aceder a trabalhos partilhados pela comunidade, mais informação sobre determinados nós e aceder a novas extensões (KNIME, 2021a).

Berthold (2006) refere-se ao KNIME como uma ferramenta exemplo considerando os requisitos de um ambiente de análise de dados modular destacando a facilidade que por um lado permite a realização de mudanças rápidas e intuitivas e por outro possibilita a exploração visual dos resultados obtidos. Também é mencionado o facto de cada nó armazenar os seus resultados permanentemente o que facilita a interrupção do fluxo de trabalho em qualquer nó e a retoma

mais tarde. Os resultados intermédios podem assim ser inspecionados a qualquer momento e novos nós podem ser inseridos sem que os nós anteriores tenham de ser reexecutados.

Segundo Berthold (2006) esta ferramenta distingue-se de outros sistemas de gestão de fluxo de trabalho como a *Mobylye* (Néron et al., 2009), *Galaxy* (Goecks et al., 2010), *Taverna* (Hull et al., 2006), *Kepler* (Ludäscher et al., 2006), *geWorkbench* (Floratos, Smith, Ji, Watkinson, & Califano, 2010), *Conveyor* (Linke, Giegerich, Goesmann, & Bateman, 2011) e muitos outros por não ser uma solução específica para um determinado domínio, integrando um forte pré-processamento de dados e capacidades analíticas. Ferramentas como a *Galaxy* e a *Taverna*, que se destinam às várias áreas das ciências da vida, tem como principal objetivo a combinação de ferramentas externas que trocam ficheiros entre si. Nativamente, a KNIME segue um caminho diferente ao encorajar uma profunda integração de ferramentas que é compatível com o formato estável da KNIME. Com esta abordagem, os dados são incorporados em células de tabela, permitindo uma fácil comunicação de ferramentas sem a necessidade de converter de ficheiros.

Uma ferramenta muito similar ao KNIME é o *Orange* que apresenta uma interface de utilizador muito intuitiva. Esta tem menos nós do que o KNIME, o que pode ser uma vantagem para iniciantes que facilmente podem obter uma visão geral das capacidades da ferramenta e consequentemente obter os *widgets* que são necessários mais facilmente. Por outro lado, os *widgets* assim que inseridos no fluxo de trabalho são imediatamente executados, permitindo uma análise mais rápida e útil para pequenas bases de dados (Demšar et al., 2013).

Relativamente às suas aplicações, este tem sido frequentemente utilizado em projetos de gestão das relações com os clientes e no setor financeiro. O foco tem se mantido no fornecimento de funcionalidades relacionadas com a construção de relatórios, análises estatísticas e o uso de linguagens de programação como o *Perl*, *Python*, *Matlab*, *R* e *Java* que permitem aos utilizadores obterem resultados orientados à suas necessidades específicas de forma mais rápida (Berthold et al., 2006).

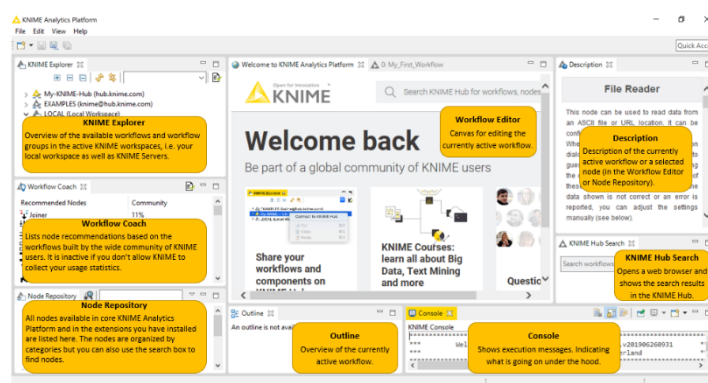


Figura 5 - Interface da plataforma KNIME (Fonte: (Knome, 2019))

3. MÉTODOS E APLICAÇÃO

Neste capítulo pretende-se descrever o método de estudo utilizado no projeto, assim como as técnicas e procedimentos aplicados. Tal como referenciado anteriormente, o CRISP-DM é a metodologia mais indicada para projetos de *data mining*, pelo que foi a escolhida para a execução do projeto. Dentro desta metodologia existem diferentes etapas e neste capítulo resume-se cada uma das mesmas.

No final do capítulo faz-se referência ao planeamento do projeto, demonstrado através de um diagrama *Ganttproject*.

3.1. Procedimento Aplicado

Com o objetivo de prever o mercado de ações para um horizonte temporal de curto prazo, utilizando para isso modelos de *machine learning*, decidiu-se seguir a metodologia CRISP-DM, bastante utilizada em projetos de *data mining*. Começou-se pela construção da base de dados, tendo o cuidado de recorrer a sites considerados fidedignos para a comunidade científica de forma a obter informação credível. Para esse feito foram integradas bases de dados provenientes do *Kaggle* com avaliação superior a bronze, onde foram revistas todas as críticas construtivas deixadas pela comunidade, e do FRED (*Federal Reserve Bank of St.Louis*). Este estudo será focado nas ações inseridas no índice *SP500* desde o dia 29/04/2013 até 07/02/2018 (1202 dias úteis), por ser um dos principais e mais influentes índices do mundo perante os investidores. Todas estas bases de dados foram posteriormente integradas no KNIME de forma a proceder-se à análise e previsão. Através do *Apêndice J*, é possível visualizar toda a lógica de integração da base de dados, através do software KNIME.

Primeiro começou-se por elaborar o *Business Understanding* onde foi possível descrever todos os objetivos que se pretendiam alcançar com o projeto, assim como algumas preocupações inerentes ao processo de previsão do mercado de ações.

De seguida, procedeu-se à realização do *Data Understanding*, onde foram descritas todas as variáveis envolvidas no *dataset* e avaliadas as suas dispersões de dados numa primeira fase. Numa segunda fase, através de um *GroupBy* foi realizado a média dos valores de fecho das ações de determinados setores pertencentes ao índice *SP500*, que permitiu estudar a performance dos vários setores e escolher as ações com maior volume para possíveis candidatos a serem previstos posteriormente. Numa última fase foram feitas as devidas correlações entre os vários atributos do *dataset* e o índice *SP500*, os vários índices de setor e a ações anteriormente escolhidas pertencente a cada um dos setores.

No final do *Data Understanding* foram tecidas as conclusões relativamente às questões analisadas, onde foram proferidas as escolhas das ações que posteriormente se iria proceder à sua previsão. O tipo de volatilidade, alta ou baixa, e de performance, “*Bear*”, “*Bull*” e “*-*”, foram os critérios aplicados. Desta forma, foram analisadas as ações das empresas Facebook, Ford Motor Company, Coca-Cola Company, Kinder Morgan, Apple, Freeport McMoRan e Host Hotels & Resorts, assim como também o índice *SP500*. Nesta fase também foram tecidas algumas hipóteses que viriam a ser concluídas na análise dos resultados das previsões.

Posteriormente procedeu-se ao tratamento e seleção de dados através do *Data Preparation*. Esta fase foi realizado a análise dos dados em falta para os vários atributos. De realçar que muita da edição e preparação dos dados foi realizada durante a criação do *dataset*, o que reduziu muito do trabalho esperado nesta fase.

De seguida, procedeu-se o *Modeling*, fase em que se realizaram as previsões das ações considerando um *dataset* constituído pelos atributos mais correlacionáveis para cada ativo estudado. Também nesta fase procedeu-se à realização da previsão do índice *SP500* de forma a comprovar se este permitiria obter melhores resultados que uma ação individual. Quanto ao critério de correlação utilizado, foi a presença de uma correlação muito forte com o índice do setor e forte e muito forte com a ação em específico e, para o índice *SP500*, variáveis com correlação forte e muito forte com o próprio índice. Como modelos de previsão, foram usados o *Linear Regression*, o *Random Forest Regression* e o *MultiLayer Perceptron*. De referir que nesta fase foram realizadas as devidas análises de sensibilidade aos parâmetros dos modelos. Em termos de horizontes temporais analisadas, foram realizadas análises aos resultados provenientes quer da utilização exclusiva do dia anterior, quer da utilização dos três dias anteriores.

Para finalizar, foram realizadas as devidas conclusões de resultados, através da fase *Evaluation*. Para esta foi criado um modelo de persistência em que o valor da previsão para o dia seguinte é o valor da ação no dia anterior. Caso haja um modelo apresente um *Root Mean Squared Error* menor que o modelo de referência, considerar-se-á um modelo que atinge o objetivo pretendido. Quanto mais significativa for essa diferença, mais considerável poderá ser o modelo.

Através da *Figura 6* é possível constatar o planeamento do projeto, com recurso a um diagrama do *Ganttproject*.

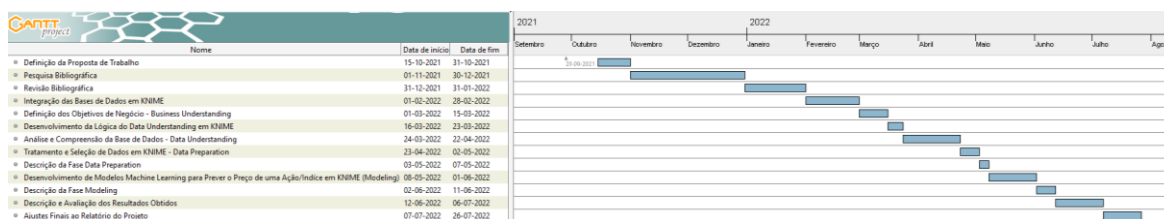


Figura 6 - Planeamento do Projeto

4. RESULTADOS E DISCUSSÃO

Neste capítulo procede-se à realização do estudo pretendido, seguindo para isso os passos da metodologia CRISP-DM. Cada passo será descrito toda a análise e serão discutidos todos os resultados obtidos.

O primeiro passo é o *Business Understanding* onde se procedeu a uma descrição dos principais objetivos do negócio e do projeto. De seguida procede-se o *Data Understanding*, onde é realizado uma descrição dos atributos que constituem a base de dados, assim como a exploração das possíveis correlações entre elas com o objetivo de obter uma melhor compreensão dos dados. Após um conhecimento mais detalhado dos atributos passa-se ao tratamento e seleção dos dados que se verificaram anteriormente serem mais relevantes para a previsão, através do *Data Preparation*. Já na fase de previsão, são desenvolvidos vários modelos de *machine learning* na fase *Modeling* e por fim, todos estes modelos são testados e validados na fase *Evaluation*.

4.1. Business Understanding

Este estudo tem como finalidade a formulação de conhecimento procurando padrões de dados que suportem hipóteses, usando para isso uma variada gama de métodos e técnicas de *data mining*. Baseado nestes padrões, pretende-se construir modelos de previsão através do conceito de *machine learning*, que permitirão conjetar com um determinado grau de incerteza o valor de uma determinada ação.

Sendo este um processo que depende intimamente dos dados recolhidos, há fatores que podem colocar a precisão do estudo em causa, nomeadamente o tratamento de dados entre as várias correlações, bem como todas as variáveis envolvidas. Do ponto de vista do estudo, um resultado bem sucedido seria um em que seja possível tecer conclusões com base nas previsões, assim como através das correlações entre as variáveis de estudo.

4.1.1. Determinação dos Objetivos de Negócio

Os principais objetivos a atingir com esta análise podem ser descritos nos seguintes pontos:

- Prever para o dia seguinte, com base no dia anterior, o valor da cotação do índice *SP500*;
- Prever para o dia seguinte, com base no dia anterior, o valor da cotação de uma determinada ação de cada setor pertencente ao *SP500*;
- Prever para o dia seguinte, com base nos últimos três dias, o valor da cotação do índice *SP500*;
- Prever para o dia seguinte, com base nos últimos três dias, o valor da cotação de uma determinada ação de cada setor pertencente ao *SP500*;
- Analisar a relação entre o preço de fecho do índice *SP500* com os pares de moeda *AUD/USD*, *EUR/USD*, *NZD/USD*, *GBP/USD*, *CAD/USD*, *CNY/USD*, *JPY/USD* e o *CHF/USD*;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com os pares de moeda *AUD/USD*, *EUR/USD*, *NZD/USD*, *GBP/USD*, *CAD/USD*, *CNY/USD*, *JPY/USD* e o *CHF/USD*;

- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com os pares de moeda *AUD/USD*, *EUR/USD*, *NZD/USD*, *GBP/USD*, *CAD/USD*, *CNY/USD*, *JPY/USD* e o *CHF/USD*;
- Analisar a relação entre o preço de fecho do índice *SP500* com o preço de fecho das três principais criptomoedas do período de amostra, nomeadamente a *Bitcoin*, a *Ethereum* e a *Litecoin*;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com o preço de fecho das três principais criptomoedas do período de amostra, nomeadamente a *Bitcoin*, a *Ethereum* e a *Litecoin*;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com o preço de fecho das três principais criptomoedas do período de amostra, nomeadamente a *Bitcoin*, a *Ethereum* e a *Litecoin*;
- Analisar a relação entre o preço de fecho do índice *SP500* com o preço de fecho das principais *commodities*, nomeadamente o ouro, o petróleo bruto, a prata e o cobre;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com o preço de fecho das principais *commodities*, nomeadamente o ouro, o petróleo bruto, a prata e o cobre;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com o preço de fecho das principais *commodities*, nomeadamente o ouro, o petróleo bruto, a prata e o cobre;
- Analisar a relação entre o preço de fecho do índice *SP500* com o preço de fecho dos principais índices de mercado dos Estados Unidos, nomeadamente o *SP500*, o *NASDAQ* e o *DJIA*;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com o preço de fecho dos principais índices de mercado dos Estados Unidos, nomeadamente o *SP500*, o *NASDAQ* e o *DJIA*;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com o preço de fecho dos principais índices de mercado dos Estados Unidos, nomeadamente o *SP500*, o *NASDAQ* e o *DJIA*;
- Analisar a relação entre o preço de fecho do índice *SP500* com o índice que representa a diferença entre o vencimento constante das obrigações de tesouro de dez anos e o vencimento constante das obrigações de tesouro de dois anos dos Estados Unidos;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com o índice que representa a diferença entre o vencimento constante das obrigações de tesouro de dez anos e o vencimento constante das obrigações de tesouro de dois anos dos Estados Unidos;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com o índice que representa a diferença entre o vencimento constante das obrigações de tesouro de dez anos e o vencimento constante das obrigações de tesouro de dois anos dos Estados Unidos;
- Analisar a relação entre o preço de fecho do índice *SP500* com a taxa de desemprego dos Estados Unidos;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com a taxa de desemprego dos Estados Unidos;

- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com a taxa de desemprego dos Estados Unidos;
- Analisar a relação entre o preço de fecho do índice *SP500* com o índice *S&P CoreLogic Case-Shiller U.S. National Home Price NSA*, que mede a variação no valor do mercado de habitação residencial nos Estados Unidos, através do acompanhamento dos preços de compra de casas unifamiliares;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com o índice *S&P CoreLogic Case-Shiller U.S. National Home Price NSA*, que mede a variação no valor do mercado de habitação residencial nos Estados Unidos, através do acompanhamento dos preços de compra de casas unifamiliares;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com o índice *S&P CoreLogic Case-Shiller U.S. National Home Price NSA*, que mede a variação no valor do mercado de habitação residencial nos Estados Unidos, através do acompanhamento dos preços de compra de casas unifamiliares;
- Analisar a relação entre o preço de fecho do índice *SP500* com a taxa de inflação dos Estados Unidos;
- Analisar a relação entre o preço de fecho dos vários índices de setor pertencentes ao *SP500* com a taxa de inflação dos Estados Unidos;
- Analisar a relação entre o preço de fecho de uma ação pertencente a um determinado setor do índice *SP500* com a taxa de inflação dos Estados Unidos.

Relativamente aos fatores que podem influenciar o resultado do projeto, estes podem ser descritos nos seguintes pontos:

- Qualidade das bases de dados utilizadas;
- Não inclusão de dados/fatores fundamentais para determinar o comportamento do preço de uma determinada ação;
- Eficácia do modelo de previsão utilizado;
- Falta de recursos que evitarão o desenvolvimento de uma análise mais aprofundada (KNIME);
- Qualidade da ferramenta/modelo de previsão criado.

Já quanto aos critérios estabelecidos para atingir os resultados pretendidos para o projeto do ponto de vista do negócio, estes podem ser descritos nos seguintes pontos:

- Ter uma base de dados fidedigna;
- Encontrar modelos capazes de prever o mercado de ações com uma eficácia superior ao modelo de persistência.

4.1.2. Objetivos do Data Mining

Os principais objetivos do projeto podem ser descritos nos seguintes pontos:

- Compreensão inicial de todas as bases de dados recolhidas de forma a conseguir integrar apenas a informação pretendida;
- Elaboração de análises estatísticas e criação de gráficos, de forma a responder aos objetivos traçados;
- Tratamento e seleção dos dados;

- Desenvolvimento de modelos baseados em *machine learning* de previsão de comportamentos do preço de uma determinada ação no software KNIME, que consigam prever o futuro do preço da ação através de uma base temporal de três e um dia.

4.2. Data Understanding

Esta segunda fase da metodologia CRISP-DM baseia-se em perceber o conjunto de dados integrados, assim como as suas possíveis correlações. Uma vez que os dados são recolhidos, é necessário conhecê-los, fazer uma descrição básica dos dados e uma exploração mais detalhada das possíveis relações entre os atributos. Numa última fase, é necessário verificar a qualidade dos dados. Através do *Apêndice K*, é possível visualizar toda a lógica que permitiu concluir as análises realizadas nesta fase, através do software KNIME.

4.2.1. Descrição dos Atributos

A base de dados construída tem os seguintes fatores/atributos para análise:

- 1) *stocks_ticker: tickers* de uma ação, que permitem identificar uma ação mais facilmente, sendo que são um atributo exclusivo de cada empresa listada no *SP500*. Esta variável tem um atributo categórico nominal;
- 2) *stocks_name*: Nome da ação, que retrata o nome oficial da empresa. Esta variável tem um atributo categórico nominal;
- 3) *stock_setor*: Setor no qual a empresa se insere. Esta variável tem um atributo categórico nominal;

No *SP500* existem onze setores, que funcionam como um *benchmarking* para as empresas ao qual abrangem, pois nestas estão incluídas todas as empresas com missão de negócio semelhante. Os setores são:

- *Communication Services* – Todos os serviços de comunicação, desde o telefone à internet de alta velocidade, que nos permitem manter conectados. Existem grandes empresas neste setor, sendo que a *AT&T* e a *Verizon* são uma das mais populares. O setor é assim constituído pelas indústrias de serviços de telecomunicações diversificados, serviços de telecomunicações sem fio, entretenimento, meios de comunicação e serviços iterativos.
- *Communication Discretionary* – Todos os produtos consumidos no dia a dia pela sociedade, desde eletrodomésticos, carros, roupa, telemóvel, entre outros. Empresas como a *Apple*, a *Disney* e a *Starbucks* são uma das principais neste setor. Normalmente este é afetado pela procura, o que o faz ser mais dependente do fator recessão e das condições económicas da sociedade. Algumas dessas indústrias são, a indústria automóvel e seus componentes, a de distribuidores, a de serviços ao consumidor, a de restauração e lazer, de equipamentos domésticos, de vestuário e de internet e marketing direto.
- *Consumer Staples* – Todos os produtos que são necessários ao ser humano, utilizados para higiene pessoal e comida. Algumas das empresas mais conhecidas neste setor são a *Procter & Gamble* e a *Kroger*. O setor é assim constituído pelas indústrias de bebidas, de retalho, de produtos alimentares, de produtos domésticos, produtos pessoais e de tabaco.

- *Energy* – Consiste em empresas que extraem e refinam matérias-primas tais como o petróleo e o gás. Empresas como a *Exxon Mobil*, a *Chevron* e a *Halliburton* são algumas das principais. Neste setor estão incluídas as indústrias de equipamento energético e serviços e as indústrias do óleo, gás e combustíveis consumíveis.
- *Financials* – Consiste em bancos, companhias de seguro, fundos de investimento imobiliário hipotecário (REITs), emitentes de cartão de crédito, entre outros. Empresas como a *JPMorgan Chase*, *Goldman Sachs* e o *Bank of America* são exemplos de ações que se enquadram neste setor. Fazem parte deste setor as indústrias da banca, mercados de capitais, financiamento ao consumidor, de serviços financeiros, seguros, fundos de investimento imobiliário hipotecário (REITs) e de finanças hipotecárias ou relacionadas com empréstimos.
- *Health care* – Consiste em todas as empresas de medicamentos, fornecimento médico e outras atividades científicas que se preocupem com a melhoria da qualidade e vida do ser humano em termos de saúde. Alguns dos principais exemplos são a *Johnson & Johnson*, a *Gilead* e a *Pfizer*. Neste setor estão inseridos as indústrias da biotecnologia, de equipamentos e fornecimento relacionado com os cuidados de saúde, de serviços e prestação de cuidados de saúde, de tecnologia relacionada com cuidados de saúde, de ferramentas e serviços de ciências da vida e farmacêutica.
- *Industrials* – Este setor engloba empresas desde caminhos de ferro, conglomerados industriais até armas militares. As principais empresas neste setor são a *Lockheed Martin* e a *CSX*. O setor divide-se assim nas indústrias aeroespacial e de defesa, de transporte aéreo e logística, de produtos de construção, de serviços comerciais e de fornecimento, de construção e engenharia, de equipamentos elétricos, de conglomerados industriais, de máquinas, da marinha, de serviços profissionais, rodoviária e ferroviária, de empresas de negociação e distribuidores e de infraestruturas.
- *Information technology* – Consiste em todas as atividades relacionadas com *hardware*, *software*, equipamento informático e serviços de TI. As principais empresas neste setor são a *Alphabet* e a *Microsoft*. Este setor apresenta as indústrias de equipamentos de comunicações, de equipamentos eletrônicos, instrumentos e componentes, de serviços de TI, de semicondutores e equipamentos relacionados e de *software/hardware*.
- *Materials* – Empresas que fornecem outros setores com matérias-primas de que necessitam para realizar os seus negócios, desde materiais preciosos, papel e produtos químicos até contentores de transporte, madeira e minério industrial. As principais empresas são o *Dow*, *Dupont*, *Ecolab* e *International Paper*. Neste estão contidos as indústrias química, de materiais de construção, de contentores e embalagens, de metais e mineração e de produtos de papel e floresta.
- *Real Estate* – Inclui todos os fundos de investimento imobiliário (REITs), com exceção dos REITs hipotecários, e empresas que gerem e desenvolvem propriedades. Empresas como a *American Tower*, *Simon Property Group* e a *Prologis* são populares neste setor. Este setor divide-se nas indústrias de investimento imobiliário de capital e de gestão imobiliária e desenvolvimento.
- *Utilities* – Estas empresas permite o funcionamento de luz, água, eletricidade, entre outros, para a sociedade em geral. Empresas como a *Exelon* e a *Dominion Resources* são uma das mais significantes. Dentro do setor encontram-se as indústrias de fornecimento elétrico, de fornecimento de gás, de produtores de energia e eletricidade renovável.

- 4) *stocks_date*: A data correspondente ao dia da amostra. Esta vai desde maio de 2013 até fevereiro de 2018, o que dá aproximadamente cerca de cinco anos de amostra. Esta variável é numérica e continua.
- 5) *stocks_open*: O preço de abertura da respetiva ação num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% das ações apresenta preços de abertura abaixo de 63,45 dólares.
- 6) *stocks_high*: O preço mais alto a que a respetiva ação num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% das stocks apresenta preços diários mais altos abaixo de 64 dólares.
- 7) *stocks_low*: O preço mais baixo a que a respetiva ação num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% das stocks apresenta preços diários mais baixos abaixo de 62,9 dólares.
- 8) *stocks_close*: O preço de fecho da respetiva ação num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% das stocks apresenta preços de fecho abaixo de 63,46 dólares.
- 9) *stocks_volume*: O volume que a ação obteve num determinado dia. Esta variável é numérica e discreta.
- 10) *AUD/USD*: O preço de fecho do dólar australiano em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *AUD/USD* apresenta valores abaixo de 1,29 dólares.
- 11) *EUR/USD*: O preço de fecho do euro em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *EUR/USD* apresenta valores abaixo de 0,88 dólares.
- 12) *NZD/USD*: O preço de fecho do dólar da nova Zelândia em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *NZD/USD* apresenta valores abaixo de 1,37 dólares.
- 13) *GBP/USD*: O preço de fecho do *pound* inglês em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *GBP/USD* apresenta valores abaixo de 0,66 dólares.
- 14) *CAD/USD*: O preço de fecho do dólar canadiano em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *CAD/USD* apresenta valores abaixo de 1,26 dólares.
- 15) *CNY/USD*: O preço de fecho do yuan chinês em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *CNY/USD* apresenta valores abaixo de 6,36 dólares.
- 16) *JPY/USD*: O preço de fecho do iene japonês em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *JPY/USD* apresenta valores abaixo de 110,74 dólares.
- 17) *CHF/USD*: O preço de fecho do franco suíço em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do par *CHF/USD* apresenta valores abaixo de 0,96 dólares.
- 18) *BTC*: O preço de fecho da cripto moeda *Bitcoin* em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço da *BTC* apresenta valores abaixo de 477,75 dólares.

- 19) *LTC*: O preço de fecho da cripto moeda *Litecoin* em dólares americanos num determinado dia. Esta variável é numérica e continua. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço da *LTC* apresenta valores abaixo de 3,96 dólares.
- 20) *ETH*: O preço de fecho da cripto moeda *Ethereum* em dólares americanos num determinado dia. Esta variável é numérica e continua. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço da *ETH* apresenta valores abaixo de 13,11 dólares.
- 21) *close_gold*: O preço de fecho do ouro em dólares americanos num determinado dia. Esta variável é numérica e continua. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do ouro apresenta valores abaixo de 1253,80 dólares.
- 22) *close_brent_crude_oil*: O preço de fecho do barril de petróleo bruto em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do barril de petróleo apresenta valores abaixo de 56,61 dólares.
- 23) *close_silver*: O preço de fecho da prata em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço da prata apresenta valores abaixo de 17,18 dólares.
- 24) *close_copper*: O preço de fecho do cobre em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do cobre apresenta valores abaixo de 2,75 dólares.
- 25) *SP500*: O preço de fecho do índice *SP500* em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do *SP500* apresenta valores abaixo de 2068,76 dólares.
- 26) *DJIA*: O preço de fecho do índice *DJIA* em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do *DJIA* apresenta valores abaixo de 17720,98 dólares.
- 27) *NASDAQ*: O preço de fecho do índice *NASDAQ* em dólares americanos num determinado dia. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o preço do *NASDAQ* apresenta valores abaixo de 4893,29 dólares.
- 28) *T10Y2Y*: O valor do índice *T10Y2Y* num determinado dia, que é atualizado todos os meses do ano. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o índice *T10Y2Y* apresenta valores abaixo de 1,37 dólares.
- 29) *FED*: O valor das taxas de juro impostas pela FED, que são atualizadas todos os meses do ano. Esta variável é numérica e percentual. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise as taxas de juro do FED apresentam valores abaixo de 0,13%.
- 30) *unemployment_rate*: A taxa de desemprego dos estados unidos, que é atualizada todos os meses do ano. Esta variável é numérica e percentual. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise a taxa de desemprego apresenta valores abaixo de 5%.
- 31) *CSUSHPINSA*: O valor do índice *S&P CoreLogic Case-Shiller U.S. National Home Price NSA*, que é atualizado todos os meses do ano. Esta variável é numérica e continua. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise o índice *CSUSHPINSA* apresenta valores abaixo de 175.06.

32) *Inflation*: O valor da inflação, que é atualizado todos os anos. Esta variável é numérica e contínua. Esta variável é numérica e percentual. Através da *Tabela 1* e do *Apêndice A*, verifica-se que 50% dos dias em análise a inflação apresenta valores abaixo de 1,46%.

Uma vez estabelecido o significado e a classificação de cada atributo, calcularam-se algumas estatísticas básicas no KNIME, encontrando-se as mesmas ilustradas na *Tabela 1*.

Tabela 1 - Estatísticas dos Atributos

ATRIBUTO	MÍNIMO	MÉDIA	MÁXIMO	DESVIO-PADRÃO	SKEWNESS	KURTOSIS
Stocks_open	1,62	82,67	1477,39	83,03	5,16	40,17
Stocks_high	1,69	83,42	1498	83,74	5,16	40,08
Stocks_low	1,61	81,92	1 450,04	82,28	5,17	40,26
Stocks_close	1,62	82,69	1 450,89	83,04	5,16	40,18
Stocks_volume	0,00	4 477 965,97	6,18E8	8 934 556,56	9,87	203,72
AUD/USD	0,96	1,24	1,46	0,12	-0,46	-1,16
EUR/USD	0,72	0,85	0,96	0,075	-0,40	-1,33
NZD/USD	1,13	1,35	1,60	0,12	-0,16	-1,02
GBP/USD	0,58	0,69	0,83	0,07	0,38	-1,20
CAN/USD	1,00	1,22	1,46	0,11	-0,41	-1,22
CNY/USD	6,04	6,41	6,96	0,27	0,44	-1,19
JPY/USD	94,29	110,15	125,58	7,87	0,09	-1,17
CHF/USD	0,85	0,96	1,03	0,04	-0,42	-0,74
BTC	68,43	1 430,50	19 114,20	2 869,46	3,73	14,62
LTC	1,16	19,12	358,34	44,36	4,56	23,16
ETH	1,00	155,78	1 299,74	256,59	2,31	5,31
Gold	1 050,80	1 247,58	1 473,90	76,79	-0,25	0,04
Brent_crude_oil	27,88	27,88	116,61	26,09	0,61	-1,24
Silver	13,67	17,78	24,65	2,30	0,60	-0,23
Copper	1,94	2,75	3,47	0,43	-0,24	-1,28
SP500	1 573,09	2 088,62	2 872,87	268,22	0,49	-0,07
DJIA	14 659,56	18 247,24	26 616,71	2 495,35	1,14	0,83
NASDAQ	3 299,13	4 987,24	7 505,77	908,68	0,54	-0,15
T10Y2Y	0,50	1,46	2,66	0,57	0,40	-1,06
FED	0,07	0,39	1,42	0,40	1,22	0,12
Unemployment_rate	4,00	5,42	7,60	0,98	0,64	-0,67
CSUSHPINSA	150,97	175,18	196,94	12,19	0,18	-1,10
Inflation	0,12	1,34	2,44	0,70	-0,69	-0,60

Com ajuda do nó “*Statistics*” foram calculados para todos os atributos os valores de mínimo, máximo, média, desvio-padrão e medidas de forma, como *skewness* e *kurtosis*.

De realçar que todos os valores referentes às ações são altamente assimétricos, pois apresentam valores de *skewness* superiores a 1, e com distribuições de dados mais estreitas que a distribuição normal, pois apresentam valores de *kurtosis* bastante superiores a 0.

Já para os pares de moedas apresentados, verifica-se uma *skewness* entre -1/2 e 1/2, o que comprova que os valores são aproximadamente simétricos. Para estes, a *kurtosis* revela-se negativa

o que demonstra uma distribuição mais achatada do que a distribuição normal, embora ambos os valores são muito próximos de 0.

Para as criptomoedas verificam-se valores de *skewness* superiores a um, o que demonstra que os dados são altamente assimétricos. Para estes, a *kurtosis* revela-se positiva o que demonstra uma distribuição mais estreita do que a distribuição normal.

Relativamente às *commodities*, verifica-se que o ouro e o cobre apresentam valores de *skewness* entre $-1/2$ e $1/2$, sendo assim os valores de cada variável aproximadamente simétricos. Já para a prata e o barril de petróleo, verifica-se uma *skewness* entre $1/2$ e 1 , sendo assim os valores de cada variável moderadamente assimétricos. Quanto ao valor de *kurtosis*, o ouro revela-se muito semelhante a um comportamento de distribuição normal e os restantes apresentam valores negativos, ou seja, tem uma distribuição dos dados mais achatada que a distribuição normal.

Quanto aos índices de mercado, o *SP500* apresenta valores de *skewness* entre $-1/2$ e $1/2$, sendo assim considerado aproximadamente simétrico. O *NASDAQ* apesar de muito próximo de ser considerado simétrico, apresenta um valor entre $1/2$ e 1 , pelo que é considerada moderadamente assimétrica. O *DJIA* apresenta valores superiores a um, logo é considerado altamente assimétrico. Já para a *kurtosis* o *SP500* é muito próximo do comportamento de uma distribuição normal, o *DJIA* uma distribuição mais estreita que a distribuição normal e o *NASDAQ* uma distribuição mais achatada que a distribuição normal.

Quanto aos índices e taxas analisados, tanto o *T10Y2Y* como o *CSUSHPINSA* apresentam valores de *skewness* entre $-1/2$ e $1/2$, sendo assim aproximadamente simétricos. Já a taxa de desemprego e a inflação pelos valores apresentados são considerados moderadamente assimétricos. As taxas de juro do FED são altamente assimétricas. Já para a *kurtosis* as taxas de juro do FED são o mais aproximado de uma distribuição normal, mais ainda assim são mais estreitas. Todos os restantes são mais achatadas que a distribuição normal.

4.2.2. Análise Setorial do SP500

Nesta etapa, foi efetuada uma exploração dos dados relativos ao atributo "*stock_close*", para analisar a performance dos vários setores e das várias ações que os constituem. Numa primeira fase foi efetuado um "*GroupBy*" que permitiu agrupar o preço de fecho do respetivo setor, dado pela média de todas as ações que lhe constituem. Numa segunda fase foi realizado um "*GroupBy*" para obter a performance individual de cada ação, assim como o seu volume médio diário.

Como seria impossível estudar todas as ações do *SP500*, procedeu-se ao critério de volume para assim analisar-se apenas a ação com maior volume do setor. A escolha deste critério baseou-se no facto de as ações com menor volume serem facilmente manipuladas por grandes grupos económicos, e que por essa razão poderão influenciar os resultados obtidos.

1) Communication Services

Através da *Figura 7* é possível constatar uma tendência crescente do setor, que começou em 04/2013 com um preço muito próximo de 68 dólares e acabou a 02/2018 muito próximo de 170 dólares. Estes valores representam uma valorização de aproximadamente 150%. Tal como se pode verificar na *Figura 7* a ação com maior volume médio diário nesta amostra foi a *Facebook*, pelo que será a escolhida para uma posterior análise.

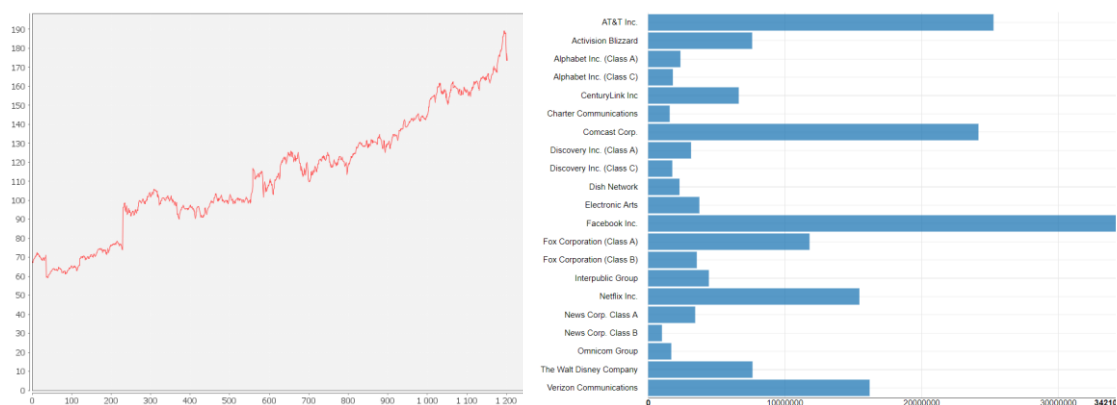
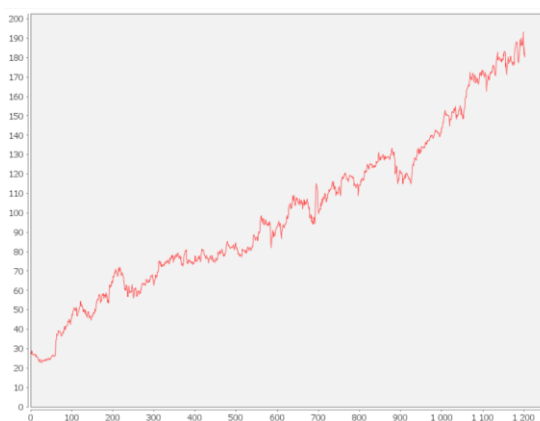


Figura 7 - Performance e Volume do setor *Communication Services*

Na *Tabela 2* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Facebook* apresenta a 3ª melhor performance do setor com 567,83% e um volume médio diário de 34.210.008,38. Na *Figura 8* é possível constatar a performance da empresa durante o período de análise.

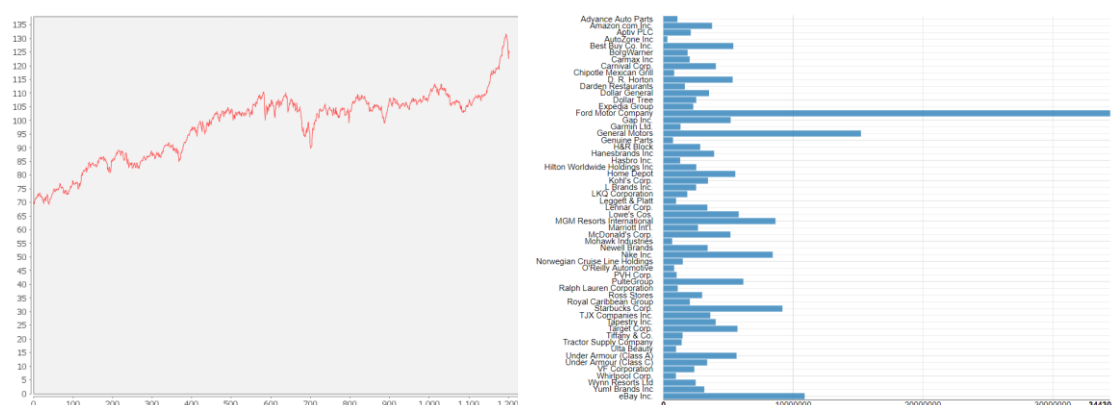
Tabela 2 - Performance e Volume médio das empresas do setor *Communication Services*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Netflix Inc.	761,31	15.448.235,48
Electronic Arts	592,46	3.745.104,44
Facebook Inc.	567,83	34.210.008,38
Activision Blizzard	365,24	7.606.905,05
Charter Communications	257,75	1.575.478,84
Alphabet Inc. (Class A)	157,46	2.361.365,32
Comcast Corp.	93,54	24.153.686,42
Alphabet Inc. (Class C)	87,76	1.808.424,17
The Walt Disney Company	66,29	7.632.425,60
Interpublic Group	54,80	4.440.465,63
Fox Corporation (Class A)	27,20	11.802.419,93
Omnicom Group	26,33	1.693.467,16
Fox Corporation (Class B)	25,20	3.558.283,56
Dish Network	11,19	2.292.816,57
News Corp. Class B	4,96	1.010.157,81
News Corp. Class A	1,65	3.440.216,94
AT&T Inc.	-0,99	25.254.187,54
Verizon Communications	-4,58	16.195.986,44
CenturyLink Inc	-54,45	6.624.803,12
Discovery Inc. (Class C)	-69,03	1.766.833,50
Discovery Inc. (Class A)	-70,67	3.138.096,44

Figura 8 - Performance da ação da empresa *Facebook*

2) Consumer Discretionary

Através da *Figura 9* é possível constatar uma tendência crescente do setor, que começou em 04/2013 com um preço muito próximo de 70 dólares e acabou a 02/2018 muito próximo de 120 dólares. Estes valores representam uma valorização de aproximadamente 71%. Tal como se pode verificar na *Figura 9* a ação com maior volume médio diário nesta amostra foi a *Ford Motor Company*, pelo que será a escolhida para uma posterior análise.

Figura 9 - Performance e Volume do setor *Consumer Discretionary*

Na *Tabela 3* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Ford Motor Company* apresenta a 54ª melhor performance do setor com -21,23% e um volume médio diário de 34.430.466,69. Na *Figura 10* é possível constatar a performance da empresa durante o período de análise.

Tabela 3a - Performance e Volume médio das empresas do setor *Consumer Discretionary*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Amazon.com Inc.	467,30	3.754.125,39
Royal Caribbean Group	254,64	2.042.890,80
Marriott Int'l.	235,51	2.670.478,96
Best Buy Co. Inc.	194,46	5.390.666,64
Lowe's Cos.	161,23	5.807.501,57
Home Depot	159,66	5.543.609,86

Tabela 3b - Performance e Volume médio das empresas do setor *Consumer Discretionary*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
MGM Resorts International	152,40	8.636.992,56
Ulta Beauty	151,45	979.926,69
Mohawk Industries	138,22	677.889,26
Ross Stores	137,34	2.987.188,78
O'Reilly Automotive	134,08	833.866,98
Expedia Group	130,70	2.305.865,22
Dollar Tree	121,97	2.533.655,25
Hasbro Inc.	114,52	1.297.403,79
Nike Inc.	109,58	8.426.791,22
Carnival Corp.	98,51	4.045.879,62
VF Corporation	93,34	2.385.884,09
Norwegian Cruise Line Holdings	90,18	1.494.833,59
Dollar General	87,58	3.519.631,85
Darden Restaurants	86,87	1.660.552,26
AutoZone Inc	80,34	312.730,61
Starbucks Corp.	79,80	9.175.965,86
Garmin Ltd.	79,00	1.321.903,77
Hanesbrands Inc	77,56	3.913.511,48
D. R. Horton	74,55	5.340.106,67
LKQ Corporation	71,48	1.845.175,49
McDonald's Corp.	62,18	5.168.616,21
TJX Companies Inc.	56,02	3.618.236,00
Lennar Corp.	50,00	3.391.158,42
Whirlpool Corp.	47,77	969.314,16
Carmax Inc	44,95	2.029.274,77
Leggett & Platt	41,60	978.542,26
Hilton Worldwide Holdings Inc	41,41	2.536.415,12
Tiffany & Co.	41,09	1.482.105,61
PulteGroup	41,02	6.161.889,06
BorgWarner	38,92	1.871.585,88
General Motors	37,68	15.224.670,56
Kohl's Corp.	32,97	3.439.413,76
Advance Auto Parts	32,41	1.083.962,93
Genuine Parts	32,04	744.435,94
Wynn Resorts Ltd	30,57	2.490.617,38
PVH Corp.	30,49	1.024.268,83
Tractor Supply Company	25,18	1.409.970,31
Yum! Brands Inc	18,54	3.155.937,14
Newell Brands	11,72	3.412.572,29
Aptiv PLC	5,54	2.112.449,52
Target Corp.	3,70	5.712.149,13
L Brands Inc.	-1,66	2.525.975,59
Under Armour (Class A)	-5,15	5.639.884,69
H&R Block	-7,91	2.839.711,77
Gap Inc.	-13,04	5.184.335,58

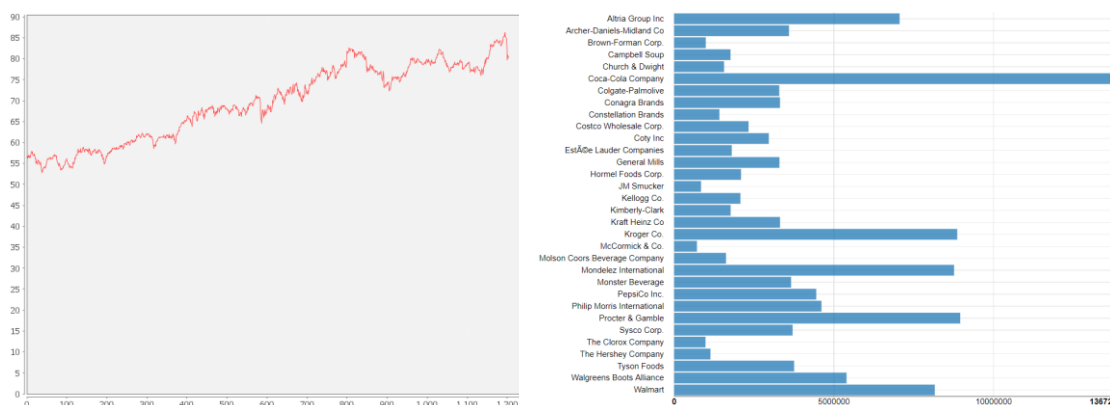
Tabela 3c - Performance e Volume médio das empresas do setor *Consumer Discretionary*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Tapestry Inc.	-14,02	4.033.178,71
eBay Inc.	-19,89	10.887.284,40
Ford Motor Company	-21,23	34.430.466,69
Chipotle Mexican Grill	-25,13	833.274,67
Ralph Lauren Corporation	-39,15	1.108.968,61
Under Armour (Class C)	-70,77	3.374.099,12

Figura 10 - Performance da empresa *Ford Motor Company*

3) Consumer Staples

Através da *Figura 11* é possível constatar uma tendência crescente do setor, que começou em 04/2013 com um preço muito próximo de 57 dólares e acabou a 02/2018 muito próximo de 79 dólares. Estes valores representam uma valorização de aproximadamente 22%. Tal como se pode verificar na *Figura 11* a ação com maior volume médio diário nesta amostra foi a *Coca-Cola Company*, pelo que será a escolhida para uma posterior análise.

Figura 11 - Performance e Volume do setor *Consumer Staples*

Na *Tabela 4* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Coca-Cola Company* apresenta a 25ª

melhor performance do setor com 5,49% e um volume médio diário de 13.672.799,46. Na *Figura 12* é possível constatar a performance da empresa durante o período de análise.

Tabela 4 - Performance e Volume das empresas do setor *Consumer Staples*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Constellation Brands	348,48	1.417.773,07
Monster Beverage	242,74	3.661.280,84
Tyson Foods	199,51	3.756.151,85
Estée Lauder Companies	95,46	1.804.400,11
Brown-Forman Corp.	84,99	991.810,61
Altria Group Inc	82,44	7.056.479,20
Costco Wholesale Corp.	68,64	2.330.151,66
Kroger Co.	67,78	8.859.096,64
Sysco Corp.	66,90	3.708.183,23
Hormel Foods Corp.	58,78	2.094.935,73
Church & Dwight	48,22	1.564.493,05
The Clorox Company	48,18	982.982,39
Molson Coors Beverage Company	47,86	1.625.305,27
Walgreens Boots Alliance	44,74	5.396.062,80
McCormick & Co.	43,30	715.654,40
Mondelez International	39,31	8.760.748,95
PepsiCo Inc.	37,33	4.449.056,89
Walmart	31,20	8.158.013,20
Archer-Daniels-Midland Co	23,98	3.595.545,67
Colgate-Palmolive	17,57	3.291.034,66
JM Smucker	17,21	843.034,21
The Hershey Company	12,39	1.137.558,94
General Mills	8,84	3.297.322,47
Kimberly-Clark	8,76	1.769.522,75
Coca-Cola Company	5,49	13.672.799,46
Procter & Gamble	5,36	8.954.241,46
Philip Morris International	3,04	4.611.555,76
Conagra Brands	1,41	3.313.696,92
Kraft Heinz Co	0,81	3.315.015,21
Coty Inc	0,69	2.964.564,98
Kellogg Co.	-1,94	2.075.776,68
Campbell Soup	-3,41	1.765.914,05

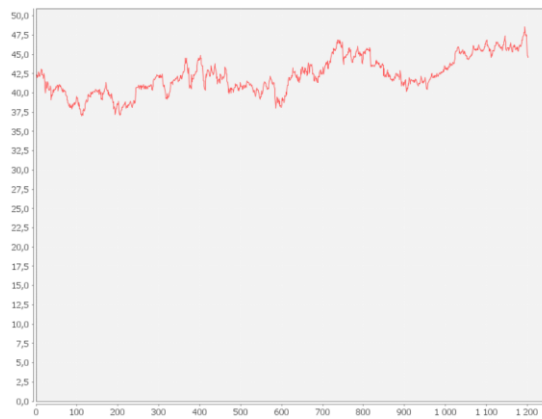


Figura 12 - Performance da empresa *Coca-Cola Company*

4) Energy

Através da *Figura 13* é possível constatar que não há tendência clara no setor, que começou em 04/2013 com um preço muito próximo de 63 dólares e acabou a 02/2018 muito próximo de 63 dólares, ainda que o preço tenha oscilado bastante durante este período. Estes valores representam uma valorização de aproximadamente 0%. Tal como se pode verificar na *Figura 13* a ação com maior volume médio diário nesta amostra foi a *Kinder Morgan*, pelo que será a escolhida para uma posterior análise.

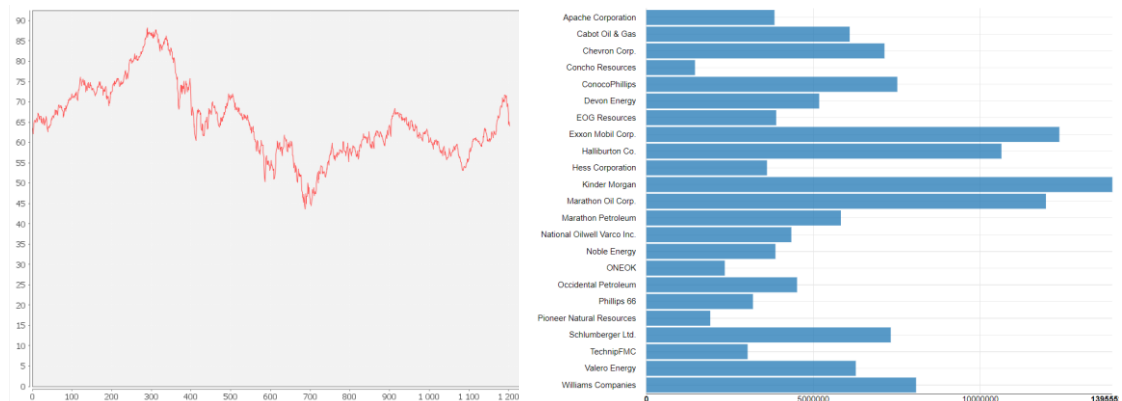
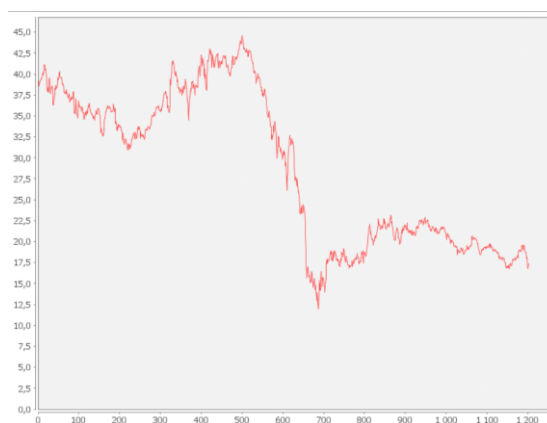


Figura 13 - Performance e Volume do setor *Energy*

Na *Tabela 5* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Kinder Morgan* apresenta a pior performance do setor com -55,55% e um volume médio diário de 13.955.528,13. Na *Figura 14* é possível constatar a performance da empresa durante o período de análise.

Tabela 5 - Performance e Volume das empresas do setor *Energy*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Valero Energy	118,06	6.272.627,18
EOG Resources	73,51	3.888.872,59
Concho Resources	71,89	1.457.960,85
Marathon Petroleum	57,97	5.825.888,58
Phillips 66	54,71	3.193.398,92
Pioneer Natural Resources	45,61	1.913.886,18
Halliburton Co.	19,53	10.634.525,57
ONEOK	10,53	2.349.622,85
Chevron Corp.	-4,97	7.131.533,34
Schlumberger Ltd.	-5,78	7.319.213,67
ConocoPhillips	-7,13	7.516.026,20
Exxon Mobil Corp.	-12,78	12.368.921,14
Occidental Petroleum	-20,48	4.513.797,17
Williams Companies	-21,51	8.077.759,76
Cabot Oil & Gas	-31,45	6.089.933,29
Devon Energy	-32,88	5.176.497,56
Hess Corporation	-38,91	3.613.373,91
TechnipFMC	-42,62	3.033.439,38
Apache Corporation	-45,94	3.837.911,85
National Oilwell Varco Inc.	-48,35	4.342.687,55
Marathon Oil Corp.	-49,91	11.966.607,00
Noble Energy	-52,09	3.865.991,97
Kinder Morgan	- 55,55	13.955.528,13

Figura 14 - Performance da empresa *Kinder Morgan*

5) Financials

Através da *Figura 15* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 51 dólares e acabou a 02/2018 muito próximo de 96 dólares. Estes valores representam uma valorização de aproximadamente 88%. Tal como se pode verificar na *Figura 15* a ação com maior volume médio diário nesta amostra foi a *Bank of America*, pelo que será a escolhida para uma posterior análise.

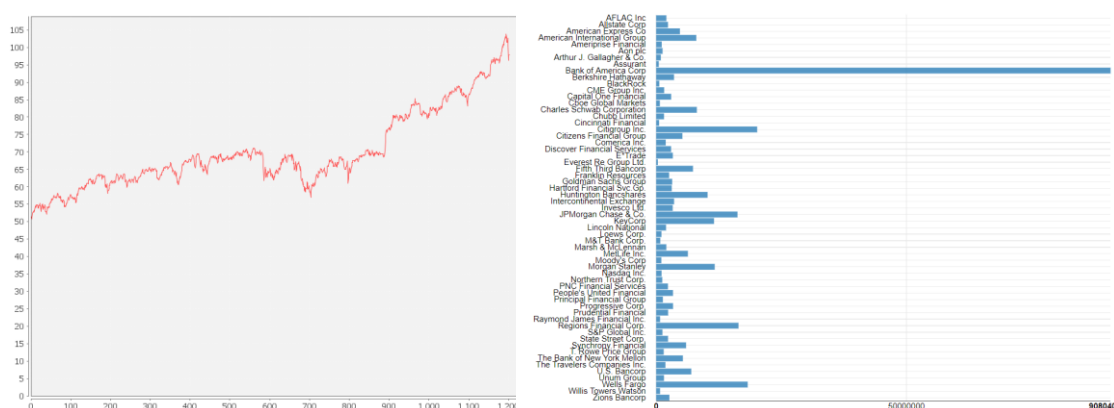


Figura 15 - Performance e Volume do setor *Financials*

Na *Tabela 6* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Bank of America* apresenta a 9ª melhor performance do setor com 152,42% e um volume médio diário de 90.804.038,44. Na *Figura 16* é possível constatar a performance da empresa durante o período de análise.

Tabela 6a - Performance e Volume das empresas do setor *Financials*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
E*Trade	400,00	3.337.099,91
S&P Global Inc.	241,80	1.260.645,33
Cboe Global Markets	209,30	741.478,10
Charles Schwab Corporation	207,14	8.134.524,20
Nasdaq Inc.	165,53	1.086.252,17
Moody's Corp	164,67	1.041.799,04
CME Group Inc.	162,87	1.604.518,44
Comerica Inc.	158,29	1.918.049,74
Bank of America Corp	152,42	90.804.038,44
Morgan Stanley	145,61	11.704.043,83
JPMorgan Chase & Co.	139,72	16.256.781,357
Aon plc	135,55	1.283.831,19
Lincoln National	133,38	1.980.568,94
PNC Financial Services	126,58	2.363.103,83
Raymond James Financial Inc.	126,27	789.222,42
Regions Financial Corp.	125,18	16.454.671,52
Zions Bancorp	120,49	2.645.895,79
Huntington Bancshares	119,36	10.276.880,40
Marsh & McLennan	118,80	2.035.753,04
Intercontinental Exchange	118,42	3.585.409,30
Ameriprise Financial	115,84	1.117.378,37
KeyCorp	113,22	11.567.230,19
Progressive Corp.	111,57	3.375.780,35
Hartford Financial Svc.Gp.	105,29	3.123.153,87
BlackRock	98,45	633.609,12
The Bank of New York Mellon	97,08	5.340.628,01
Allstate Corp	96,69	2.370.819,26

Tabela 6b - Performance e Volume das empresas do setor *Financials*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Citizens Financial Group	94,28	5.243.490,63
Fifth Third Bancorp	92,16	7.364.745,61
Northern Trust Corp.	88,85	1.225.724,50
Berkshire Hathaway	87,95	3.570.817,90
Assurant	87,68	528.640,90
M&T Bank Corp.	85,32	829.342,66
Everest Re Group Ltd.	84,64	343.419,38
Prudential Financial	84,39	2.392.696,40
Unum Group	83,73	1.568.094,91
Principal Financial Group	78,45	1.345.773,72
State Street Corp.	78,02	2.379.060,02
Goldman Sachs Group	77,18	3.202.596,39
Discover Financial Services	74,24	2.986.309,47
Capital One Financial	70,61	3.000.763,49
The Travelers Companies Inc.	66,29	1.856.944,43
Chubb Limited	65,90	1.575.481,29
U.S. Bancorp	64,61	7.018.206,67
Synchrony Financial	63,04	5.984.343,17
Citigroup Inc.	60,17	20.189.039,64
AFLAC Inc	58,40	2.046.821,99
Arthur J. Gallagher & Co.	55,81	935.441,25
Wells Fargo	51,90	18.299.003,90
Cincinnati Financial	50,40	596.534,84
T. Rowe Price Group	49,02	1.513.895,50
People's United Financial	46,57	3.367.821,54
American International Group	46,38	8.023.321,23
American Express Co	38,42	4.750.664,26
Willis Towers Watson	25,06	786.228,31
MetLife Inc.	18,16	6.342.530,54
Invesco Ltd.	12,62	3.311.704,41
Loews Corp.	7,48	1.062.504,97
Franklin Resources	-23,61	2.592.137,86

Figura 16 - Performance da empresa *Bank of America*

6) Health Care

Através da *Figura 17* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 69 dólares e acabou a 02/2018 muito próximo de 140 dólares. Estes valores representam uma valorização de aproximadamente 103%. Tal como se pode verificar na *Figura 17* a ação com maior volume médio diário nesta amostra foi a *Pfizer*, pelo que será a escolhida para uma posterior análise.

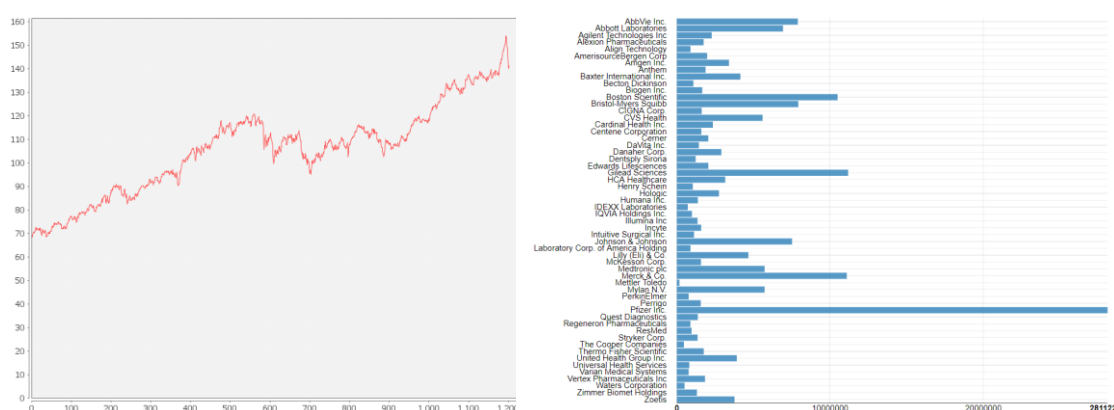


Figura 17 - Performance e Volume do setor *Health Care*

Na *Tabela 7* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Pfizer* apresenta a antepenúltima posição, sendo a 54ª melhor performance do setor com 14,95% e um volume médio diário de 90.804.038,44. Na *Figura 18* é possível constatar a performance da empresa durante o período de análise.

Tabela 7a - Performance e Volume das empresas do setor *Health Care*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Align Technology	611,60	890.177,67
Centene Corporation	347,98	1.595.901,69
Edwards Lifesciences	303,55	2.053.879,30
IDEXX Laboratories	300,25	713.440,36
Incyte	292,10	1.586.157,57
United Health Group Inc.	276,30	3.912.753,44
Humana Inc.	262,28	1.371.447,75
Boston Scientific	251,40	10.489.102,99
Illumina Inc	243,53	1.349.170,39
Anthem	227,08	1.874.158,67
Mettler Toledo	211,22	169.561,76
CIGNA Corp.	195,81	1.621.420,02
Thermo Fisher Scientific	157,00	1.755.496,01
AbbVie Inc.	151,21	7.894.818,55
PerkinElmer	147,61	772.293,50
Intuitive Surgical Inc.	145,03	1.119.844,43
HCA Healthcare	143,75	3.156.628,72

Tabela 7b - Performance e Volume das empresas do setor *Health Care*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Zoetis	136,15	3.762.046,85
Becton Dickinson	134,07	1.078.742,27
Stryker Corp.	134,04	1.353.490,52
IQVIA Holdings Inc.	133,22	983.215,75
Waters Corporation	118,48	504.082,88
The Cooper Companies	111,91	460.471,96
Vertex Pharmaceuticals Inc	98,74	1.835.914,91
Hologic	97,99	2.748.793,35
ResMed	96,25	960.254,82
Laboratory Corp. of America Holding	81,42	887.242,52
Varian Medical Systems	80,46	770.651,65
Quest Diagnostics	79,00	1.367.116,54
Universal Health Services	78,89	812.937,41
AmerisourceBergen Corp	74,29	1.975.234,87
Medtronic plc	73,68	5.726.028,70
Amgen Inc.	66,37	3.401.073,92
Agilent Technologies Inc	64,16	2.275.249,04
Henry Schein	62,26	1.040.744,08
Gilead Sciences	61,99	11.176.915,48
Abbott Laboratories	59,73	6.928.244,46
Zimmer Biomet Holdings	58,57	1.308.359,47
Danaher Corp.	58,14	2.903.562,78
Bristol-Myers Squibb	56,93	7.925.165,01
Regeneron Pharmaceuticals	55,29	883.620,17
Johnson & Johnson	53,56	7.520.015,59
Cardinal Health Inc.	48,05	2.351.465,41
McKesson Corp.	46,26	1.574.449,55
Biogen Inc.	46,19	1.657.467,32
Mylan N.V.	39,44	5.725.273,11
Dentsply Sirona	38,15	1.227.640,98
Lilly (Eli) & Co.	36,98	4.667.589,27
Cerner	33,94	2.051.617,38
CVS Health	29,44	5.595.515,69
DaVita Inc.	25,46	1.437.061,65
Alexion Pharmaceuticals	18,74	1.747.676,75
Merck & Co.	16,86	11.091.931,20
Pfizer Inc.	14,95	28.112.225,97
Baxter International Inc.	-7,10	4.144.630,75
Perrigo	-27,12	1.558.626,13

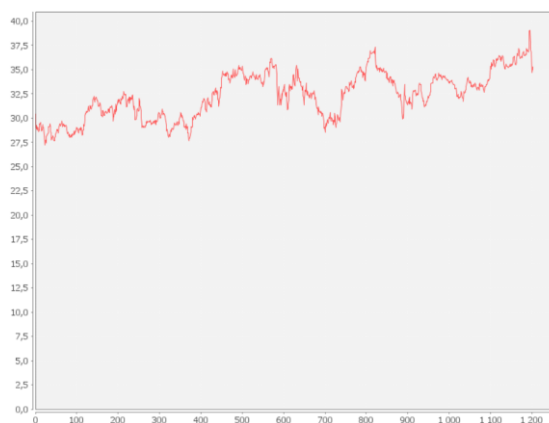


Figura 18 - Performance da empresa Pfizer

7) Industrials

Através da *Figura 19* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 63 dólares e acabou a 02/2018 muito próximo de 120 dólares. Estes valores representam uma valorização de aproximadamente 90%. Tal como se pode verificar na *Figura 19* a ação com maior volume médio diário nesta amostra foi a *General Electric*, pelo que será a escolhida para uma posterior análise.

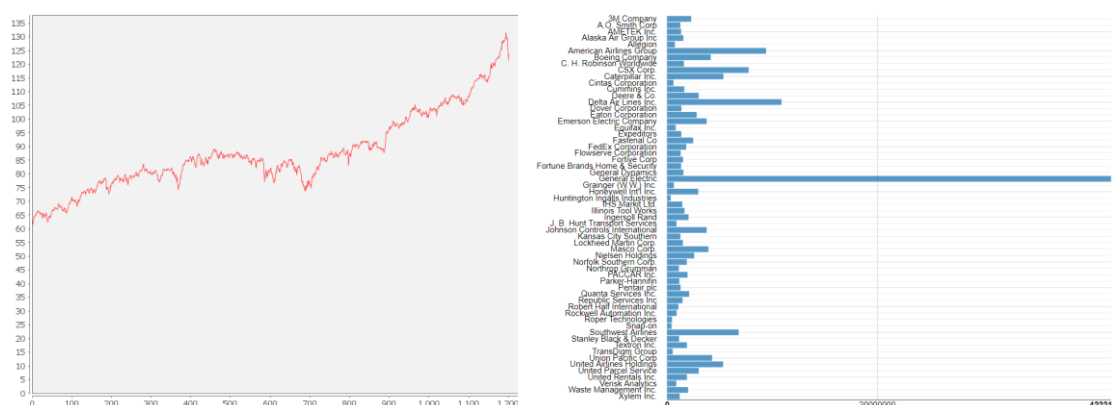


Figura 19 - Performance e Volume do setor Industrials

Na *Tabela 8* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *General Electric* apresenta a pior performance do setor com -31,42% e um volume médio diário de 42.221.564,70. Na *Figura 20* é possível constatar a performance da empresa durante o período de análise.

Tabela 8a - Performance e Volume das empresas do setor Industrials

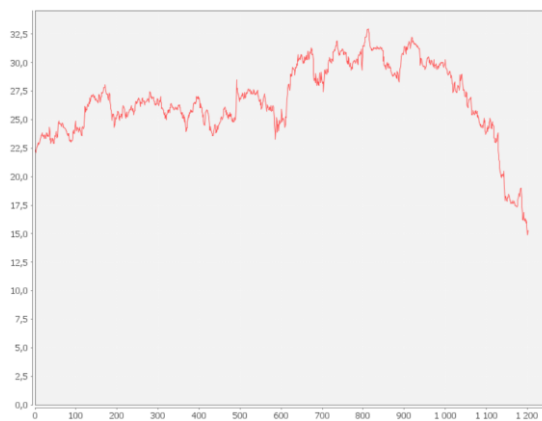
NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Northrop Grumman	350,45	1.118.078,39
Huntington Ingalls Industries	343,81	336.437,15
Southwest Airlines	327,55	6.797.129,17
Boeing Company	278,80	4.142.990,87
Cintas Corporation	249,63	603.422,42
Lockheed Martin Corp.	248,71	1.492.823,23

Tabela 8b - Performance e Volume das empresas do setor *Industrials*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
A.O. Smith Corp	233,93	1.249.085,09
United Rentals Inc.	231,20	1.880.653,95
Delta Air Lines Inc.	217,70	10.874.467,36
American Airlines Group	205,77	9.410.831,75
General Dynamics	193,33	1.542.044,34
FedEx Corporation	167,89	1.812.423,27
Xylem Inc.	154,12	1.192.100,23
Illinois Tool Works	152,11	1.655.357,24
3M Company	124,59	2.286.877,42
Roper Technologies	124,29	462.219,11
Rockwell Automation Inc.	122,49	900.026,04
Textron Inc.	121,54	1.879.232,74
CSX Corp.	115,24	7.746.988,53
Stanley Black & Decker	114,47	1.129.573,56
Alaska Air Group Inc	110,04	1.536.644,66
Masco Corp.	109,06	3.920.219,89
Parker-Hannifin	106,36	1.162.113,07
Honeywell Int'l Inc.	104,86	2.955.871,77
United Airlines Holdings	104,25	5.319.919,86
Waste Management Inc.	101,95	1.993.630,09
TransDigm Group	101,63	522.680,45
Snap-on	95,71	425.481,24
Equifax Inc.	93,04	812.621,10
Republic Services Inc	91,86	1.462.793,21
Norfolk Southern Corp.	86,27	1.867.777,92
AMETEK Inc.	82,89	1.325.765,26
Caterpillar Inc.	82,01	5.346.909,96
Deere & Co.	81,74	3.001.565,89
Union Pacific Corp	74,96	4.276.826,40
Expeditors	74,23	1.349.290,77
Fortune Brands Home & Security	73,94	1.328.309,81
Robert Half International	71,46	1.063.573,47
IHS Markit Ltd.	69,18	1.438.888,50
Ingersoll Rand	68,12	2.032.468,36
Allegion	64,95	739.441,85
J. B. Hunt Transport Services	63,30	890.687,01
C. H. Robinson Worldwide	57,33	1.590.973,83
Verisk Analytics	56,54	867.078,56
Cummins Inc.	52,53	1.629.932,46
Fortive Corp	44,71	1.512.779,91
Dover Corporation	43,36	1.352.465,44
PACCAR Inc.	37,85	1.937.915,65
Eaton Corporation	35,40	2.807.562,61
United Parcel Service	29,72	3.003.311,30
Pentair plc	25,87	1.274.663,65

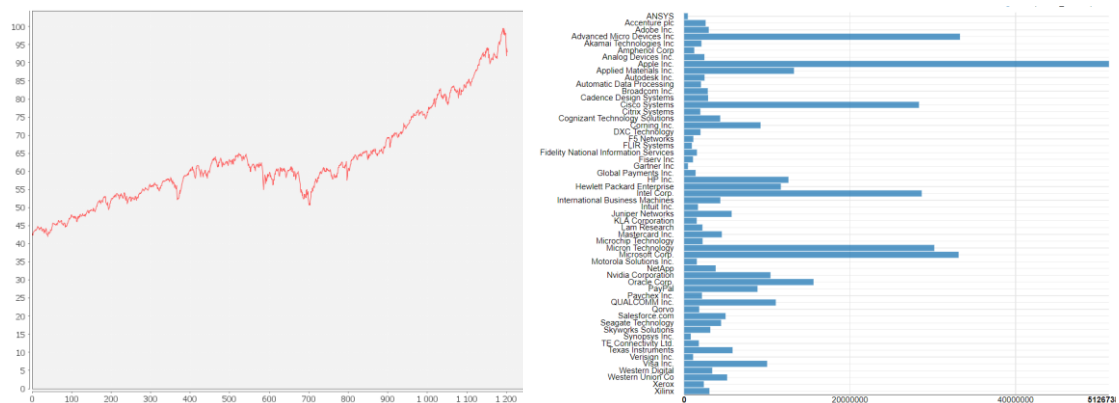
Tabela 8c - Performance e Volume das empresas do setor *Industrials*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Quanta Services Inc.	24,91	2.088.469,87
Emerson Electric Company	24,68	3.751.080,92
Johnson Controls International	11,65	3.757.957,14
Nielsen Holdings	9,73	2.575.494,36
Fastenal Co	6,65	2.469.889,02
Grainger (W.W.) Inc.	5,83	637.423,54
Kansas City Southern	-0,50	1.259.913,37
Flowserve Corporation	-21,00	1.284.653,76
General Electric	-31,52	42.221.564,70

Figura 20 - Performance da empresa *General Electric*

8) Information Technology

Através da *Figura 21* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 43 dólares e acabou a 02/2018 muito próximo de 93 dólares. Estes valores representam uma valorização de aproximadamente 116%. Tal como se pode verificar na *Figura 21* a ação com maior volume médio diário nesta amostra foi a *Apple*, pelo que será a escolhida para uma posterior análise.

Figura 21 - Performance e Volume do setor *Information Technology*

Na *Tabela 9* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Apple* apresenta a 19ª melhor performance do setor com 159,64% e um volume médio diário de 51.267.387,32. Na *Figura 22* é possível constatar a performance da empresa durante o período de análise.

Tabela 9a - Performance e Volume das empresas do setor *Information Technology*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Nvidia Corporation	1586,10	10.419.479,74
Broadcom Inc.	629,73	2.854.369,85
Global Payments Inc.	361,16	1.384.574,96
Skyworks Solutions	359,64	3.158.179,18
Micron Technology	349,31	30.186.122,05
Advanced Micro Devices Inc	332,84	33.287.514,97
Adobe Inc.	326,85	2.972.763,15
Lam Research	276,55	2.208.715,52
Applied Materials Inc.	238,13	13.256.398,37
Mastercard Inc.	213,04	4.549.092,49
Autodesk Inc.	189,30	2.464.236,18
Visa Inc.	185,10	10.017.882,27
Fiserv Inc	184,39	1.073.126,42
Texas Instruments	183,51	5.845.276,95
Cadence Design Systems	178,69	2.892.330,76
Microsoft Corp.	174,79	33.125.149,04
Salesforce.com	168,35	5.003.923,72
Intuit Inc.	164,67	1.670.454,13
Apple Inc.	159,64	51.267.387,32
Synopsys Inc.	144,52	803.062,10
Cognizant Technology Solutions	143,19	4.353.261,20
Verisign Inc.	141,20	1.084.548,53
Fidelity National Information Services	137,94	1.545.847,98
Amphenol Corp	135,73	1.223.955,70
TE Connectivity Ltd.	132,14	1.775.771,64
Microchip Technology	130,56	2.229.086,41
Gartner Inc	109,56	463.433,51
Corning Inc.	106,94	9.223.384,44
PayPal	105,91	8.854.942,96
FLIR Systems	100,41	932.703,21
ANSYS	96,56	444.676,77
Analog Devices Inc.	95,49	2.449.413,48
KLA Corporation	93,30	1.522.009,96
Cisco Systems	92,28	28.340.803,18
Accenture plc	91,24	2.593.433,47
Intel Corp.	90,24	28.672.054,18
F5 Networks	86,14	1.112.127,57
Paychex Inc.	79,19	2.150.056,63
Motorola Solutions Inc.	78,67	1.535.341,37
Xilinx	77,69	3.051.340,21

Tabela 9b - Performance e Volume das empresas do setor *Information Technology*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
HP Inc.	70,91	12.591.136,31
Automatic Data Processing	69,14	2.034.128,77
NetApp	66,90	3.807.924,35
Juniper Networks	58,10	5.732.366,75
Western Digital	54,68	3.389.939,70
Oracle Corp.	51,58	15.617.304,07
Akamai Technologies Inc	49,85	2.088.785,16
DXC Technology	41,40	1.974.555,33
Citrix Systems	40,95	1.955.527,32
Seagate Technology	36,80	4.460.667,53
Western Union Co	33,13	5.190.725,92
Qorvo	10,92	1.817.043,71
QUALCOMM Inc.	5,57	11.063.136,88
Hewlett Packard Enterprise	-5,98	11.672.718,90
Xerox	-7,75	2.378.135,08
International Business Machines	-22,75	4.367.813,28

Figura 22 - Performance da empresa *Apple*

9) Materials

Através da *Figura 23* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 60 dólares e acabou a 02/2018 muito próximo de 105 dólares. Estes valores representam uma valorização de aproximadamente 75%. Tal como se pode verificar na *Figura 23* a ação com maior volume médio diário nesta amostra foi a *Freeport-McMoRan*, pelo que será a escolhida para uma posterior análise.

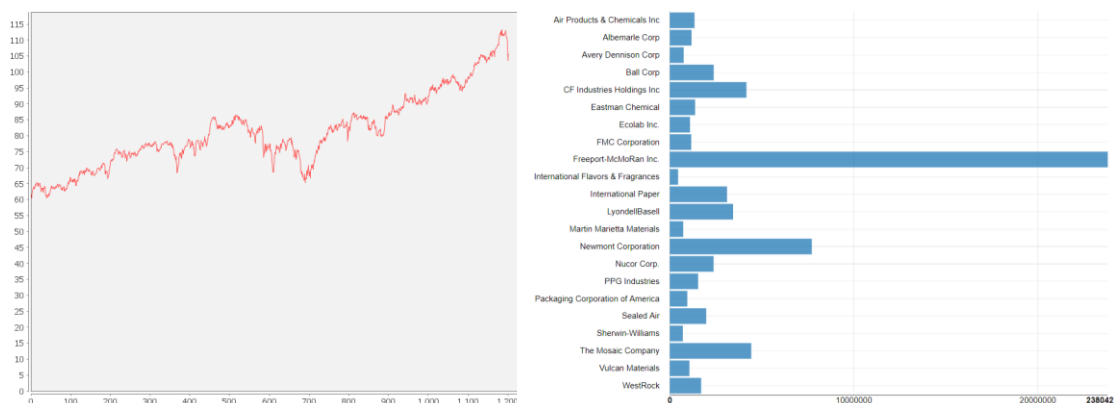


Figura 23 - Performance e Volume do setor *Materials*

Na *Tabela 10* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Freeport-McMoRan* apresenta penúltima posição em termos de performance no setor com -40,27% e um volume médio diário de 23.804.272,35. Na *Figura 24* é possível constatar a performance da empresa durante o período de análise.

Tabela 10 - Performance e Volume das empresas do setor *Materials*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
Avery Dennison Corp	175,19	753.938,80
Vulcan Materials	163,11	1.065.992,86
Packaging Corporation of America	148,53	951.499,43
Martin Marietta Materials	122,27	728.583,42
Sherwin-Williams	120,83	708.179,66
Sealed Air	100,82	1.978.527,37
International Flavors & Fragrances	86,56	449.535,55
Air Products & Chemicals Inc	82,82	1.342.131,19
LyondellBasell	79,42	3.439.622,45
Albemarle Corp	79,30	1.179.977,29
Ball Corp	76,92	2.387.234,51
Ecolab Inc.	59,90	1.098.440,87
PPG Industries	55,26	1.539.322,21
Nucor Corp.	45,85	2.379.329,49
Eastman Chemical	44,98	1.374.955,39
FMC Corporation	38,40	1.164.488,86
International Paper	25,53	3.109.583,00
Newmont Corporation	9,01	7.717.817,52
CF Industries Holdings Inc	4,56	4.164.323,78
WestRock	1,64	1.702.197,18
Freeport-McMoRan Inc.	-40,27	23.804.272,35
The Mosaic Company	-59,17	4.427.208,28



Figura 24 - Performance da empresa Freeport-McMoRan

10) Real Estate

Através da *Figura 25* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 79 dólares e acabou a 02/2018 muito próximo de 100 dólares. Estes valores representam uma valorização de aproximadamente 27%. Tal como se pode verificar na *Figura 25* a ação com maior volume médio diário nesta amostra foi a *Host Hotels & Resorts*, pelo que será a escolhida para uma posterior análise.

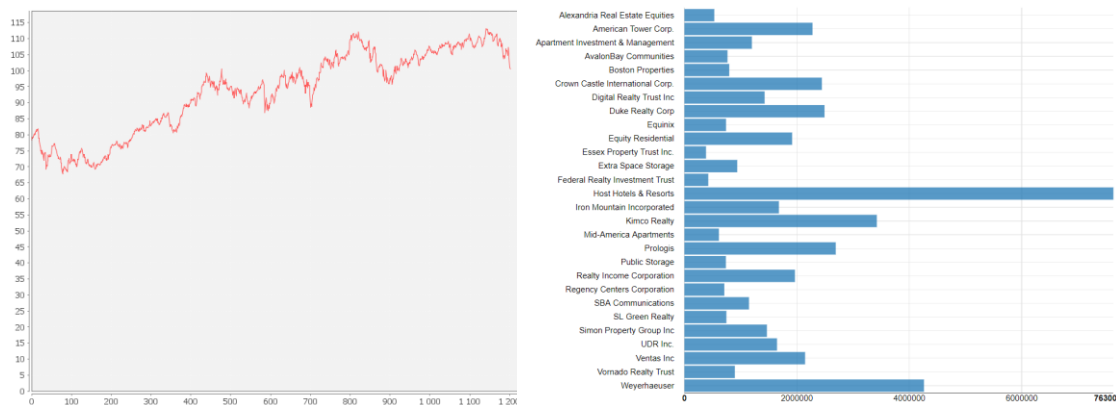


Figura 25 - Performance e Volume do setor *Real Estate*

Na *Tabela 11* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Host Hotels & Resorts* apresenta 17ª melhor performance no setor com 8,96% e um volume médio diário de 7.630.099,56. Na *Figura 26* é possível constatar a performance da empresa durante o período de análise.

Tabela 11 - Performance e Volume das empresas do setor *Real Estate*

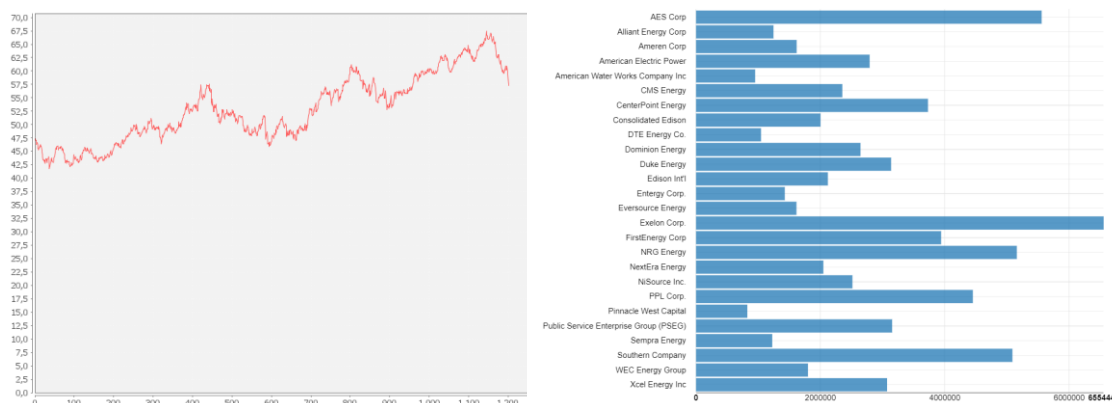
NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
SBA Communications	113,83	1.148.508,02
Equinix	101,96	740.088,36
Extra Space Storage	87,20	939.675,01
American Tower Corp.	68,59	2.278.630,76
Alexandria Real Estate Equities	66,40	531.631,58
Digital Realty Trust Inc	48,99	1.426.493,89
Duke Realty Corp	47,5	2.491.885,57
Prologis	43,91	2.691.560,25
Crown Castle International Corp.	40,60	2.444.977,53
Essex Property Trust Inc.	40,50	384.359,48
UDR Inc.	37,90	1.646.982,55
Mid-America Apartments	27,99	613.652,01
Apartment Investment & Management	26,36	1.199.849,62
AvalonBay Communities	20,78	766.424,84
Public Storage	14,42	737.826,89
Weyerhaeuser	13,44	4.260.983,00
Host Hotels & Resorts	8,96	7.630.099,56
Boston Properties	5,67	797.053,69
SL Green Realty	4,69	746.420,32
Regency Centers Corporation	4,014	709.322,59
Equity Residential	-0,612	1.915.847,98
Realty Income Corporation	-2,84	1.965.713,28
Federal Realty Investment Trust	-3,418	424.408,32
Iron Mountain Incorporated	-10,78	1.680.688,45
Simon Property Group Inc	-12,00	1.467.322,44
Vornado Realty Trust	-22,59	896.957,43
Ventas Inc	-32,95	2.146.238,71
Kimco Realty	-38,16	3.422.595,72



Figura 26 - Performance da empresa Host Hotels & Resorts

11) Utilities

Através da *Figura 27* é possível constatar uma tendência crescente no setor, que começou em 04/2013 com um preço muito próximo de 47,5 dólares e acabou a 02/2018 muito próximo de 57,5 dólares. Estes valores representam uma valorização de aproximadamente 21%. Tal como se pode verificar na *Figura 27* a ação com maior volume médio diário nesta amostra foi a *Exelon*, pelo que será a escolhida para uma posterior análise.

Figura 27 - Performance e Volume do setor *Utilities*

Na *Tabela 12* é possível ter conhecimento sobre quais as ações que obtiveram melhores performances, assim como o seu volume médio diário exato. A *Exelon* apresenta 20ª melhor performance no setor com -4,31% e um volume médio diário de 6.554.448,96. Na *Figura 28* é possível constatar a performance da empresa durante o período de análise.

Tabela 4a - Performance e Volume das empresas do setor *Utilities*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
American Water Works Company Inc	88,54	950.897,92
NextEra Energy	82,24	2.047.501,34
Ameren Corp	47,14	1.618.171,09
Alliant Energy Corp	42,05	1.245.647,61
CMS Energy	41,71	2.352.636,51
DTE Energy Co.	37,51	1.045.348,34
WEC Energy Group	34,64	1.800.110,39
Xcel Energy Inc	34,45	3.072.178,45
Public Service Enterprise Group (PSEG)	30,94	3.153.222,64
Eversource Energy	26,25	1.615.054,24
Sempra Energy	26,20	1.225.344,42
American Electric Power	25,61	2.791.695,00
Pinnacle West Capital	23,46	825.075,56
Dominion Energy	20,50	2.644.510,80
Consolidated Edison	18,28	2.001.796,00
Edison Int'l	10,58	2.119.992,80
CenterPoint Energy	7,50	3.729.693,43
Entergy Corp.	2,70	1.427.453,25
Duke Energy	-1,06	3.137.004,07
Exelon Corp.	-4,31	6.554.448,96

Tabela 12b - Performance e Volume das empresas do setor *Utilities*

NOME DA AÇÃO	PERFORMANCE (%)	VOLUME MÉDIO (QT)
PPL Corp.	-9,13	4.451.245,44
NRG Energy	-9,23	5.158.408,08
Southern Company	-9,97	5.087.265,91
AES Corp	-23,28	5.557.111,32
NiSource Inc.	-25,88	2.514.794,46
FirstEnergy Corp	-33,48	3.940.969,33

Figura 28 - Performance e Volume da empresa *Exelon*

4.2.3. Correlações Entre Atributos

Nesta fase procedeu-se à realização das correlações entre os ativos e os atributos que constituem o dataset.

1) *SP500*

Considerando que todas as ações estão inseridas no *SP500*, assumiu-se este índice como um *benchmarking* para fazer uma macro análise da correlação que existe entre a maioria das ações do *SP500* e todas as variáveis em estudo.

Durante os 1202 dias úteis de amostra, que como mostra a *Figura 29* foi um período de *bull market*, onde a maioria das empresas tiveram uma performance positiva. É possível verificar também que o índice *SP500* teve uma performance de aproximadamente 68%.

Entre o começo da amostra, no mês de abril de 2013, até aproximadamente no dia 550 de análise, cerca de meio do ano de 2015, há uma tendência positiva em ambos os índices. Já no período entre meio de 2015 e o final de 2016 é possível verificar um período em que não há uma tendência clara, havendo muito oscilação nos índices de mercado. Por fim, desde o final de 2016 até muito próximo do final da amostra, ou seja, início de 2018 verifica-se um período de tendência crescente muito vertical que leva de seguida a uma pequena correção de tendência, tal como se pode constatar no final da amostra.

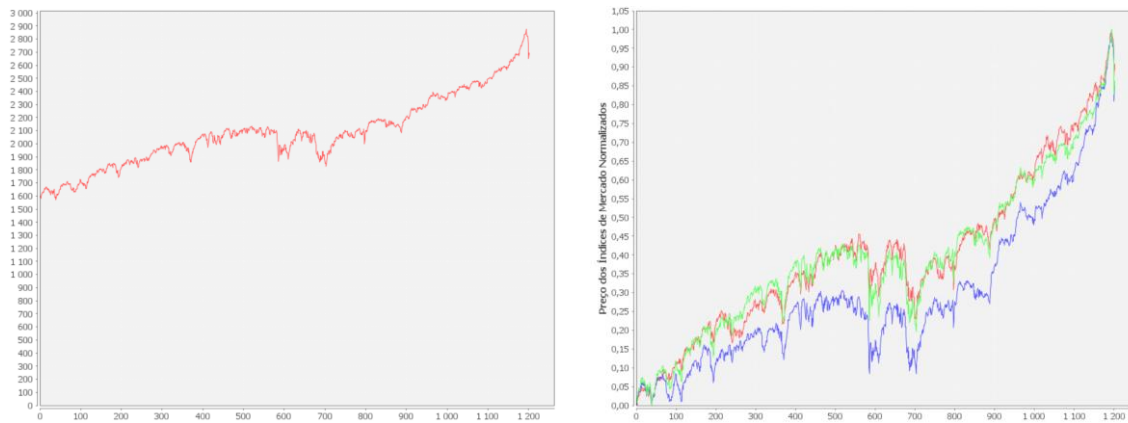


Figura 29 - Performance do SP500 e preços de fecho do SP500 (Vermelho), DJIA(Azul) e NASDAQ(Verde) normalizados

Através da Figura 30 observa-se que não existe qualquer relação entre o SP500 e as commodities ouro e cobre, pois, apresentam valores absolutos inferiores a 0,3. Relativamente aos pares de moeda NZD/USD e JPY/USD e a inflação, este apresenta uma correlação positiva fraca e com a prata uma correlação negativa fraca, tendo valores absolutos entre 0,3 e 0,5. Em relação aos pares AUD/USD, EUR/USD, CAD/USD, CNY/USD e CHF/USD, e às criptos BTC e LTC verifica-se uma correlação positiva moderada e com o preço do barril de petróleo uma correlação negativa moderada, tendo valores absolutos entre 0,5 e 0,7. Com o par GBP/USD, a crypto ETH e as taxas de juro da FED apresenta uma correlação positiva forte e com o índice T10Y2Y e a taxa de desemprego uma correlação negativa forte, com valores absolutos entre 0,7 e 0,9. Por fim com os índices de mercado DJIA e NASDAQ e o índice CSUSHPINSa apresenta uma correlação positiva muito forte, com valores absolutos superiores a 0,9.

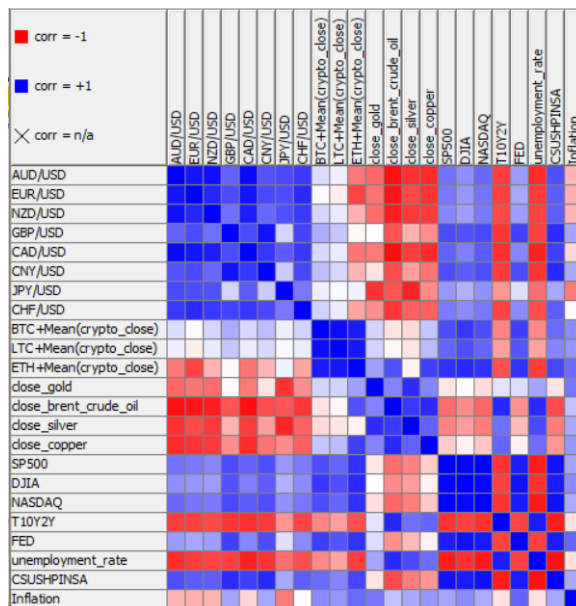


Figura 30 - Correlações com o índice SP500

De seguida, na *Figura 31 e 32*, são apresentados de uma forma visual todas as correlações fortes e muito fortes que existem com o índice *SP500*. Todos os valores são normalizados para que essa percepção seja mais evidente.



Figura 31 - Correlações Fortes com o *SP500*(Azul): *GBP/USD*(amarelo), *ETH*(Verde) e *FED*(Vermelho)



Figura 32 - Correlações Muito Fortes com o *SP500*(Azul) : *DJIA*(Lilás), *NASDAQ*(Amarelo), *T10Y2Y*(Verde), Taxa de Desemprego(Azul) e *CSUSHPINSA*(Vermelho)

2) Setor *Communication Services*

Relativamente aos pares de moedas, é possível constatar que há uma correlação positiva fraca entre o par *JPY/USD* e o índice do setor. Por outro lado, com os pares *AUD/USD*, *EUR/USD*, *NZD/USD* e *CHF/USD* verifica-se uma correlação positiva moderada. Já com os pares *GBP/USD*, *CAD/USD*, *CNY/USD* é possível visualizar uma correlação positiva forte.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva forte com o preço da *ETH*. Com a *BTC* e *LTC* apresenta uma correlação positiva moderada.

Em termos de *commodities*, é possível notar que o índice do setor não tem qualquer correlação com o ouro. Já com o cobre apresenta uma correlação negativa fraca e com o preço do barril de petróleo e a prata, este apresenta uma correlação negativa moderada.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva muito forte com os vários índices de mercado, apresentando valores superiores a 0,9.

Por fim, é possível verificar que o índice do mercado apresenta uma correlação positiva fraca com a inflação (0,37). Já com as taxas de juro do FED apresenta uma correlação positiva forte (0,89) e com o valor do índice *CSUSHPINS*A apresenta uma correlação positiva muito forte (0,98). Por outro lado, com o índice *T10Y2Y* apresenta uma correlação negativa forte (-0,87) e com o valor da taxa de desemprego apresenta uma correlação negativa muito forte (-0,94).

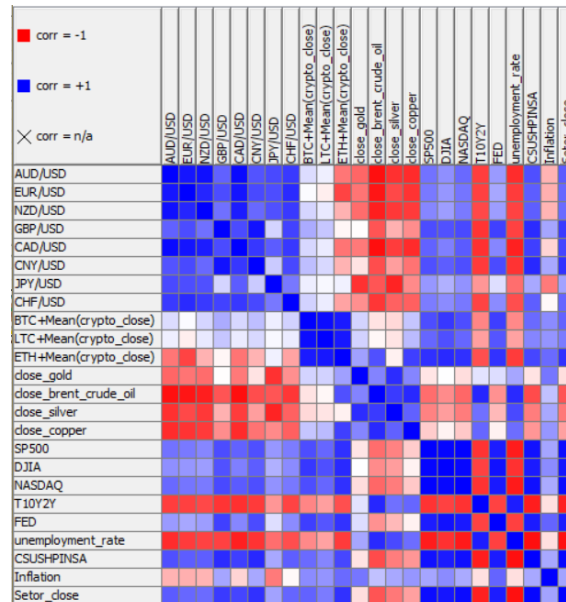


Figura 33 - Correlações com o setor *Communication Services*

– Facebook

Relativamente aos pares de moedas, é possível constatar que não há qualquer tipo de relação entre o par *JPY/USD* e o preço de fecho da *Facebook*, sendo que o valor da correlação é inferior a 0,3. Por outro lado verifica-se uma correlação positiva forte entre os pares *GBP/USD*, *CAD/USD*, *CNY/USD* e o preço de fecho da *Facebook*. Também é possível visualizar a correlação negativa fraca (-0,52) entre o volume médio diária e o preço da ação.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Facebook* tem uma correlação positiva forte com o preço da *ETH*. Com a *BTC* e *LTC* apresenta uma correlação positiva moderada.

Em termos de *commodities*, é possível notar que o preço de fecho da *Facebook* não tem qualquer correlação com o ouro e com o cobre. Já com o preço do barril de petróleo, este apresenta uma correlação negativa moderada (0,68).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Facebook* apresenta uma correlação positiva muito forte com os vários índices de mercado e com o índice do setor.

Por fim, é possível verificar que o preço de fecho da *Facebook* apresenta uma correlação positiva fraca com a inflação (0,37). Já com as taxas de juro do FED e com o valor do índice *CSUSHPINS*A apresenta uma correlação positiva muito forte. Por outro lado, com o índice *T10Y2Y* apresenta uma correlação negativa forte (-0,87) e com o valor da taxa de desemprego apresenta uma correlação negativa muito forte (-0,93).

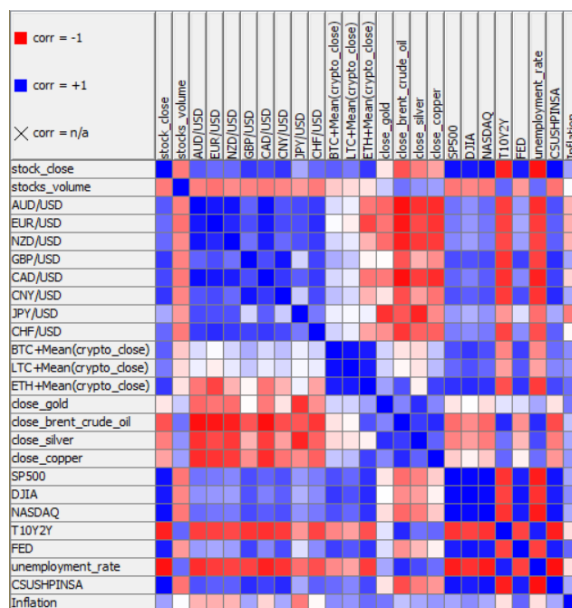


Figura 34 - Correlações com a empresa *Facebook*

3) Setor *Consumer Discretionary*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva moderada com os pares *GBP/USD*, *CNY/USD*, *JPY/USD* e *CHF/USD*. Por outro lado, com os pares *AUD/USD*, *EUR/USD*, *NZD/USD* e *CAD/USD* verifica-se uma correlação positiva forte.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva fraca com a *BTC* e a *LTC*. Já com a *ETH*, verifica-se uma correlação positiva forte (0,82).

Em termos de *commodities*, é possível notar que o índice do setor tem uma correlação negativa fraca com o ouro (0,07) e uma correlação negativa moderada com cobre. Já com o preço do barril de petróleo e com a prata, este apresenta uma correlação negativa forte.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva forte com todos os índices de mercado estudados, tendo valores entre 0,7 e 0,9.

Por fim, é possível verificar que o índice do setor não apresenta qualquer tipo de correlação com a inflação. Já com as taxas de juro do FED apresenta uma correlação positiva moderada e com o valor do índice *CSUSHPINSA* apresenta uma correlação positiva forte (0,88). Por outro lado, com o índice *T10Y2Y* apresenta uma correlação negativa forte (-0,84) e com a taxa de desemprego apresenta uma correlação negativa muito forte (-0,94).

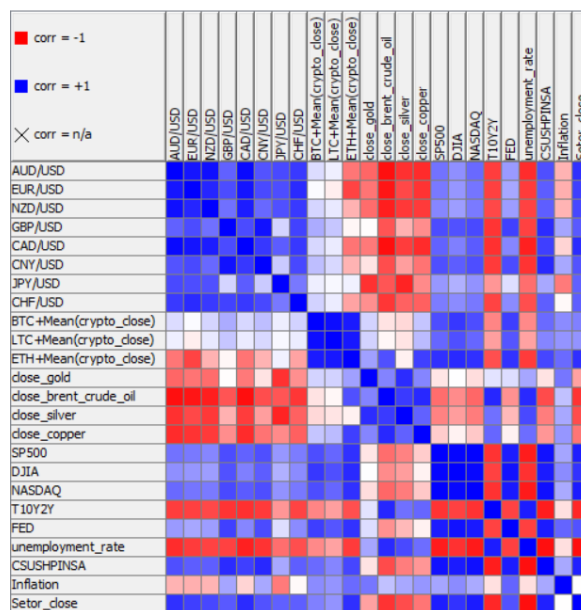


Figura 35 - Correlações com o setor *Consumer Discretionary*

– *Ford Motor Company*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Ford Motor Company* não apresenta qualquer correlação com o par *JPY/USD* e com o seu volume médio diário, sendo que o valor da correlação é inferior a 0,3. Por outro lado verifica-se uma correlação negativa forte entre os pares *GBP/USD*, *CAD/USD*, *CNY/USD* e o preço de fecho da *Ford Motor Company*.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Ford Motor Company* não tem qualquer tipo de correlação com a *ETH* e com a *LTC*. Já com a *BTC*, verifica-se uma correlação negativa fraca (-0,38).

Em termos de commodities, é possível notar que o preço de fecho da *Ford Motor Company* não tem qualquer correlação com o ouro (0,07). Já com a prata é uma correlação positiva fraca, com o cobre uma correlação positiva moderada e com o preço do barril de petróleo, este apresenta uma correlação positiva forte (0,68).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Ford Motor Company* apresenta uma correlação negativa moderada com os índices *DJIA* e o índice do setor, ambas muito próximas de serem fortes. Já com os índices *SP500* e com o *NASDAQ* é possível constatar uma correlação negativa forte.

Por fim, é possível verificar que o preço de fecho da *Ford Motor Company* apresenta uma correlação negativa fraca com a inflação (-0,35). Já com as taxas de juro do FED e com o valor do índice *CSUSHPINSA* apresenta uma correlação negativa forte. Por outro lado, com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação positiva forte.

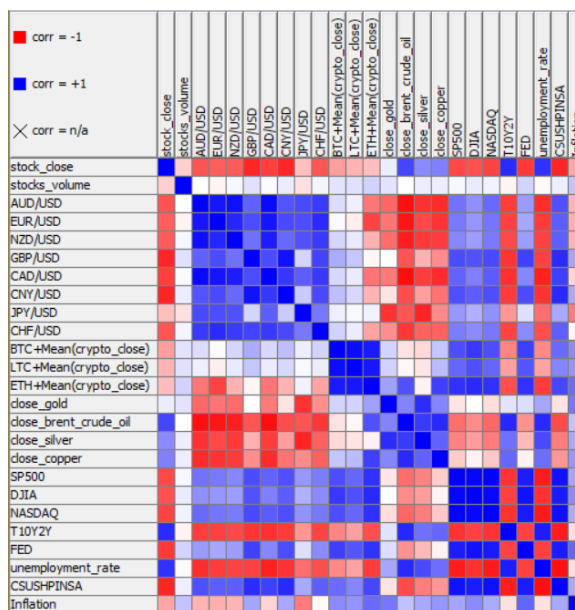


Figura 36 - Correlações com a empresa *Ford Motor Company*

4) Setor *Consumer Staples*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com o par *JPY/USD*. Por outro com todos os restantes pares lado verifica-se uma correlação positiva forte.

Quanto às criptomoedas, é possível averiguar que o índice de mercado tem uma correlação positiva fraca com a *BTC* e com a *LTC*. Já com a *ETH*, verifica-se uma correlação positiva moderada (0,58).

Em termos de *commodities*, é possível notar que o preço de fecho do índice do setor não tem qualquer correlação com o ouro. Já com a prata e o cobre, este apresenta uma correlação negativa moderada e com o preço do barril de petróleo uma correlação negativa forte (-0,83).

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva forte com todos os índices de mercado.

Por fim, é possível verificar que o preço do índice do setor não apresenta qualquer tipo de correlação com a inflação. Já com as taxas de juro do FED apresenta uma correlação positiva forte (0,77) e com o valor do índice *CSUSHPINSA* apresenta uma correlação positiva muito forte (0,94). Por outro lado, com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa muito forte.

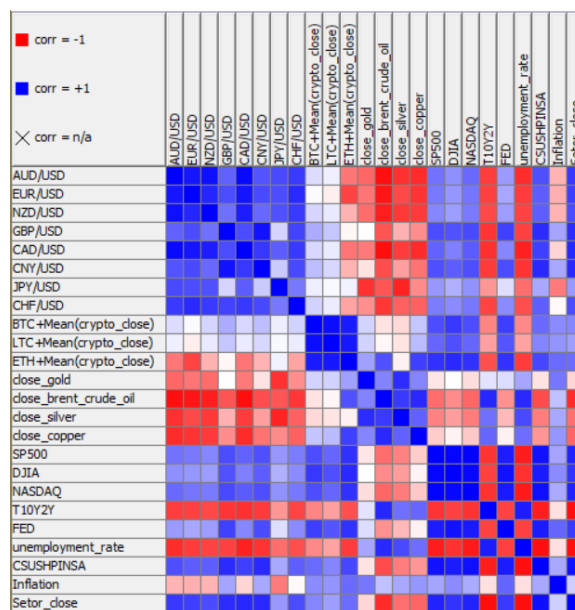


Figura 37 - Correlações com o setor *Consumer Staples*

– *Coca-Cola Company*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Coca-Cola Company* não apresenta qualquer correlação com o par *JPY/USD* e com o seu volume médio diário, sendo que o valor da correlação é inferior a 0,3. Por outro lado apenas verifica-se uma correlação positiva moderada entre os pares *GBP/USD*, *CNY/USD* e o preço de fecho da *Coca-Cola Company*.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Coca-Cola Company* tem uma correlação positiva fraca com a *LTC* (0,45). Já com a *ETH* e com a *BTC*, verifica-se uma correlação positiva moderada.

Em termos de *commodities*, é possível notar que o preço de fecho da *Coca-Cola Company* não tem qualquer correlação com o ouro (0,07). Já com o preço do barril de petróleo, este apresenta uma correlação negativa moderada (0,56) e com a prata e o cobre uma correlação negativa fraca.

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Coca-Cola Company* apresenta uma correlação positiva forte com todos os índices de mercado e o índice do setor.

Por fim, é possível verificar que o preço de fecho da *Coca-Cola Company* apresenta uma correlação positiva fraca com a inflação (0,34). Já com as taxas de juro do FED e com o valor do índice *CSUSHPINSA* apresenta uma correlação positiva forte. Por outro lado, com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte.

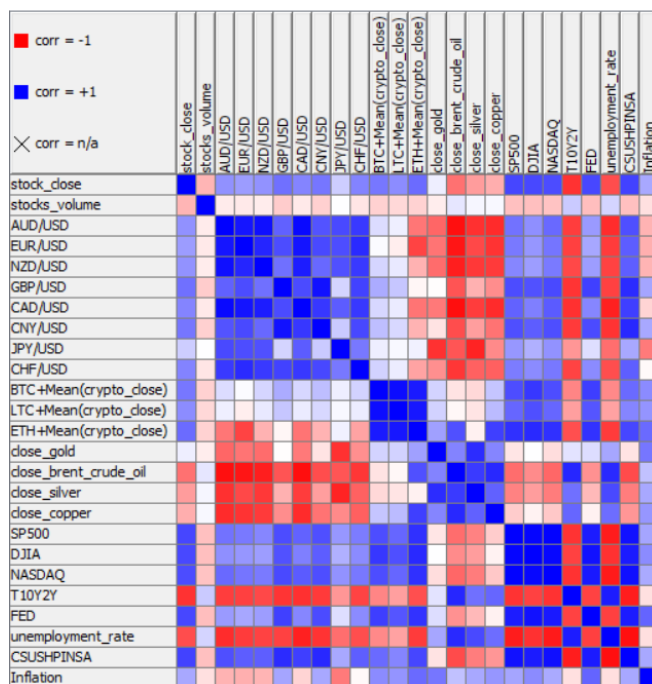


Figura 38 - Correlações com a empresa *Coca-Cola Company*

5) Setor *Energy*

Relativamente aos pares de moedas, é possível constatar que o preço do índice do setor apresenta uma correlação negativa fraca com o par *JPY/USD*. Por outro lado, verifica-se uma correlação negativa moderada com os pares *GBP/USD*, *CNY/USD* e *CHF/USD*. Já com os pares *AUD/USD*, *EUR/USD*, *NZD/USD* e *CAD/USD* apresenta uma correlação negativa forte.

Quanto às criptomoedas, é possível averiguar que o índice do setor não tem qualquer tipo de correlação com a *BTC* e com a *LTC*, mas com a *ETH* apresenta uma correlação positiva fraca (0,40).

Em termos de *commodities*, é possível notar que o índice do setor tem uma correlação positiva fraca com o ouro. Já com a prata apresenta uma correlação positiva moderada e com o preço do barril de petróleo e o cobre uma correlação positiva forte.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação negativa fraca com os índices de mercado *SP500* e *NASDAQ*. Já com o índice *DJIA* não se verifica qualquer tipo de correlação.

Por fim, é possível verificar que o índice do setor não apresenta qualquer correlação com a inflação. Já com as taxas de juro do FED este apresenta uma correlação negativa fraca e com o valor do índice *CSUSHPINSA* apresenta uma correlação negativa moderada. Por outro lado, com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação positiva moderada.

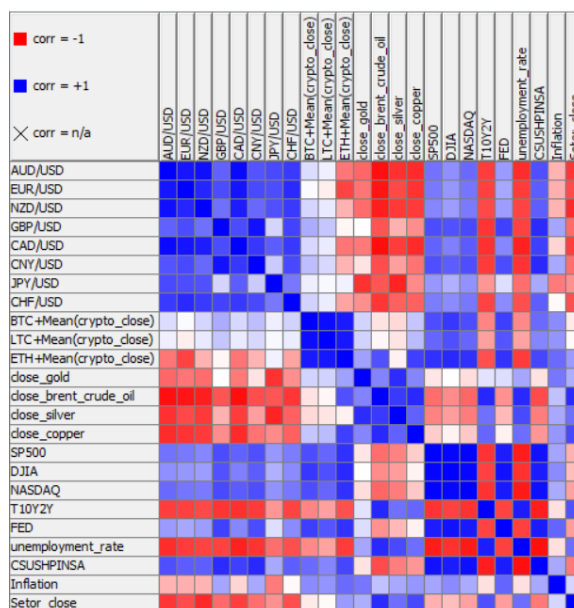


Figura 39 - Correlações com o setor *Energy*

– *Kinder Morgan*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Kinder Morgan* não apresenta qualquer correlação com o seu volume médio diário nem com o par *JPY/USD*, sendo que o valor da correlação é inferior a -0,3. Por outro lado verifica-se uma correlação negativa moderada entre o preço de fecho da *Kinder Morgan* e os pares *AUD/USD*, *EUR/USD*, *NZD/USD* e *CHF/USD* e uma correlação negativa forte com os pares *GBP/USD*, *CAD/USD* e *CNY/USD*.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Kinder Morgan* não tem qualquer tipo de correlação com a *ETH* (-0,24). Já com a *LTC* (-0,32) e com a *BTC* (-0,41) apresenta uma correlação negativa fraca.

Em termos de *commodities*, é possível notar que o preço de fecho da *Kinder Morgan* não tem qualquer relação com o ouro. Já com a prata apresenta uma correlação positiva fraca. Por outro lado, com o cobre e com o preço do barril de petróleo, este apresenta uma correlação positiva moderada.

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Kinder Morgan* apresenta uma correlação positiva moderada com o índice do setor. Já com os índices de mercado *SP500*, *DJIA* e *NASDAQ* verifica-se uma correlação negativa moderada.

Por fim, é possível verificar que o preço de fecho da *Kinder Morgan* apresenta uma correlação negativa fraca com a inflação. Já com as taxas de juro do FED e o valor do índice *CSUSHPINSA* apresenta uma correlação negativa forte. Por outro lado, com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação positiva forte.

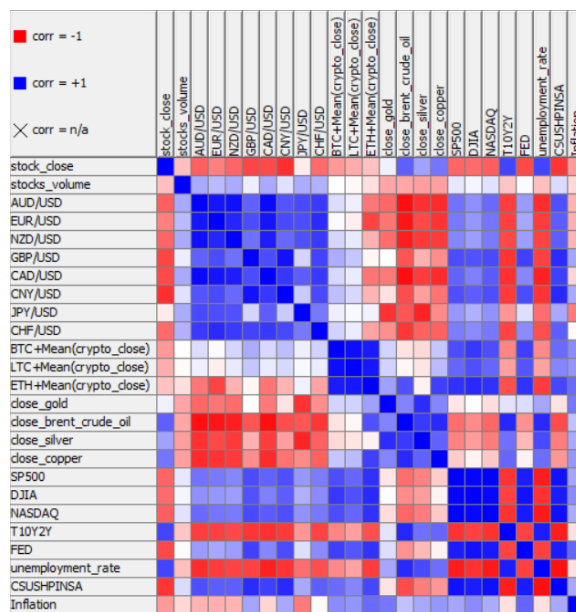


Figura 40 - Correlações com a empresa *Kinder Morgan*

6) Setor *Financials*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com os pares *AUD/USD*, *EUR/USD*, *NZD/USD*, *JPY/USD* e *CHF/USD*. Já com os pares *GBP/USD*, *CAD/USD* e *CNY/USD*, verifica-se uma correlação positiva moderada.

Quanto às criptomoedas, é possível averiguar que o índice de mercado tem uma correlação positiva moderada com a *LTC* (0,65). Já com a *BTC* e com a *ETH* é possível verificar uma correlação positiva forte.

Em termos de *commodities*, é possível notar que o índice de mercado não apresenta qualquer tipo de correlação com o ouro e com o cobre. Já com o preço do barril de petróleo e com a prata apresenta uma correlação negativa fraca.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva muito forte com todos os índices de mercado, apresentando valores superiores a 0,9.

Por fim, é possível verificar que o índice do setor apresenta uma correlação positiva fraca com a inflação. Já com a *T10Y2Y* e a taxa de desemprego verifica-se uma correlação negativa forte. Por outro lado, com o valor do índice *CSUSHPINSA* verifica-se uma correlação positiva forte e com as taxas de juro do FED uma correlação positiva muito forte.

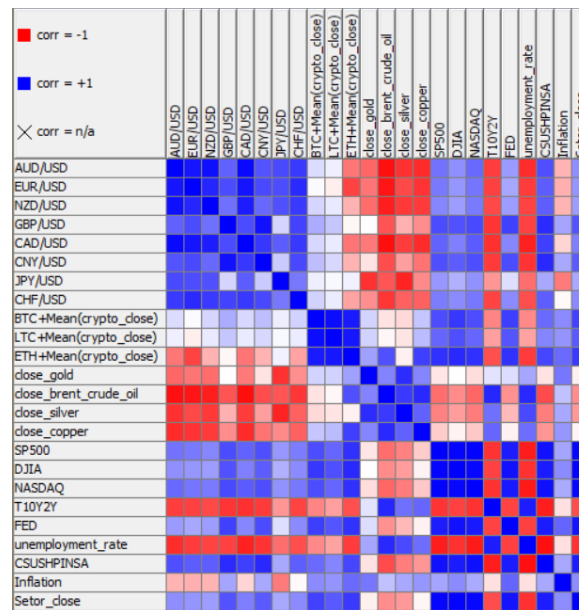


Figura 41 - Correlações com o setor *Financials*

– *Bank of America*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Bank of America* não apresenta qualquer correlação com o seu volume médio diário e os pares *AUD/USD*, *EUR/USD*, *NZD/USD*, *JPY/USD*. Por outro lado, verifica-se uma correlação positiva moderada entre o preço de fecho da *Bank of America* com os pares *GBP/USD* e *CNY/USD* e uma correlação positiva fraca para os pares *CAD/USD* e *CHF/USD*.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Bank of America* tem uma correlação positiva moderada com a *LTC* (0,68). Já com a *BTC* e com a *ETH* é possível verificar uma correlação positiva forte.

Em termos de *commodities*, é possível notar que o preço de fecho da *Bank of America* não apresenta qualquer tipo de correlação com nenhuma das *commodities* apresentadas, sendo todos os valores inferiores a 0,3 em valor absoluto.

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Bank of America* apresenta uma correlação positiva forte com os índices de mercado *SP500* e *NASDAQ* e uma correlação positiva muito forte com o índice do setor e com o *DJIA*.

Por fim, é possível verificar que o preço de fecho da *Bank of America* apresenta uma correlação positiva moderada com a inflação e uma correlação negativa moderada com a *T10Y2Y* e a taxa de desemprego. Já com as taxas de juro do *FED* e com o valor do índice *CSUSHPINSA* apresenta uma correlação positiva forte.

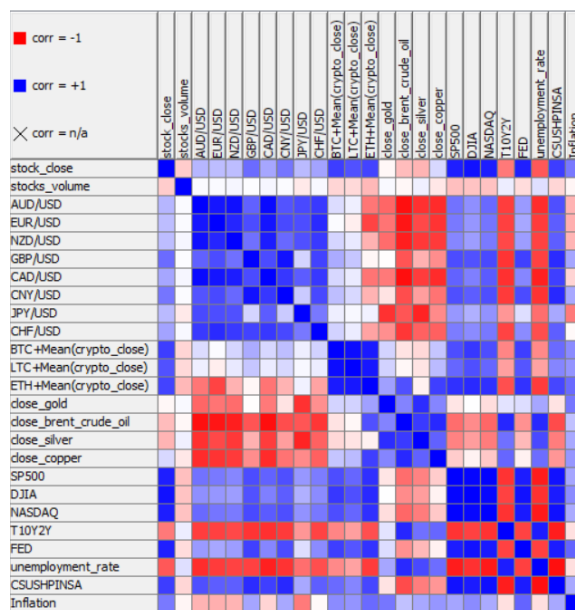


Figura 42 - Correlações com a empresa *Bank of America*

7) Setor *Health Care*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva forte com o par *CAD/USD* (0,72). Com os restantes verifica-se uma correlação positiva moderada.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva moderada com a *BTC* e a *LTC*. Já com a *ETH* apresenta uma correlação positiva forte (0,84).

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro, sendo que o valor da correlação é inferior a 0,3 em valor absoluto. Já com o cobre apresenta uma correlação negativa fraca, com a prata uma correlação negativa moderada e com o preço do barril de petróleo uma correlação negativa forte (-0,72).

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva forte com o índice de mercado *DJIA*. Já com o *SP500* e com o *NASDAQ* apresenta uma correlação positiva muito forte.

Por fim, é possível verificar que o índice do setor não apresenta qualquer tipo de correlação com a inflação. Com o índice do FED apresenta uma correlação positiva forte (0,79) e com o índice *CSUSHPINSIA* uma correlação positiva muito forte (0,91). Já com o *T10Y2Y* é uma correlação negativa forte (-0,84) e com a taxa de desemprego uma correlação negativa muito forte (-0,93).

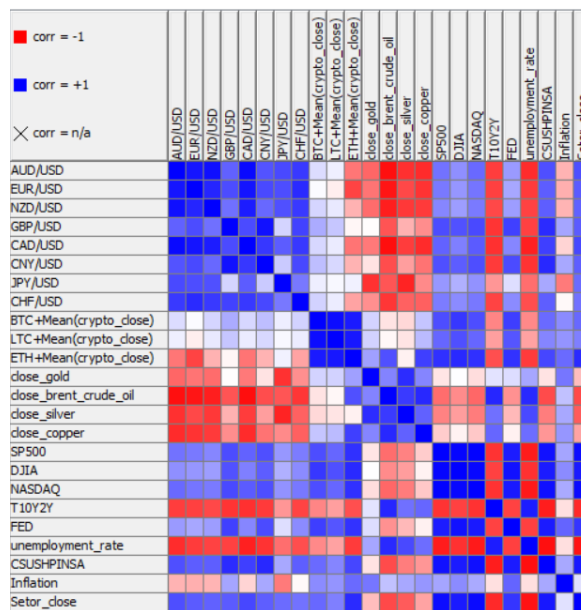


Figura 43 - Correlações com o setor *Health Care*

– Pfizer

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Pfizer* não apresenta qualquer correlação com o seu volume médio diário. Por outro lado, verifica-se uma correlação positiva fraca entre o preço de fecho da *Pfizer* com os pares *CNY/USD*, *JPY/USD* e *CHF/USD* e uma correlação positiva fraca para os pares *CAD/USD* e *CHF/USD*. Todos os restantes pares apresentam uma correlação positiva moderada.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Pfizer* tem uma correlação positiva fraca com a *LTC*, *BTC* e com a *ETH*.

Em termos de *commodities*, é possível notar que o preço de fecho da *Pfizer* não apresenta qualquer tipo de correlação com o ouro, sendo que o valor da correlação é inferior a 0,3 em valor absoluto. Já com a prata e com o cobre apresenta uma correlação negativa fraca e com o preço do barril de petróleo uma correlação negativa moderada (-0,60).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Pfizer* apresenta uma correlação positiva forte com todos os índices de mercado e com o índice do setor, apresentando valores entre 0,7 e 0,9.

Por fim, é possível verificar que o preço de fecho da *Pfizer* não apresenta qualquer tipo de correlação com a inflação. Com o índice do FED apresenta uma correlação positiva moderada e com o índice *CSUSHPINSA* uma correlação positiva forte. Já com o *T10Y2Y* é uma correlação negativa moderada e com a taxa de desemprego uma correlação negativa forte.

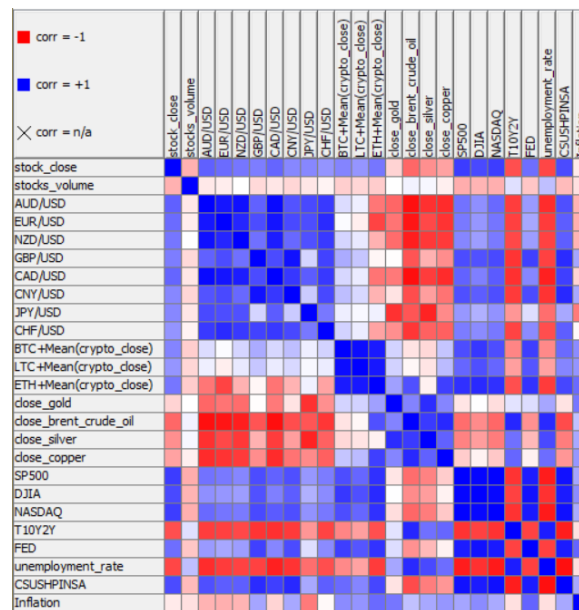


Figura 44 - Correlações com a empresa *Pfizer*

8) Setor *Industrials*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com os pares *AUD/USD*, *EUR/USD*, *NZD/USD*, *JPY/USD* e *CHF/USD*. Por outro lado, verifica-se uma correlação positiva moderada com os pares *CAD/USD* e *CNY/USD* e uma correlação forte com o par *GBP/USD* (0,70).

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva moderada com *LTC*. Já com a *BTC* e a *ETH* esta apresenta uma correlação positiva forte.

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro e com o cobre, sendo que o valor da correlação é inferior a 0,3 em valor absoluto. Já com a prata apresenta uma correlação negativa fraca e com o preço do barril de petróleo apresenta uma correlação negativa moderada (0,50).

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva muito forte com todos os índices de mercado estudados.

Por fim, é possível verificar que o índice do setor apresenta uma correlação positiva fraca com a inflação. Já com as taxas de juro do FED e com o índice *CSUSHPINSA* apresenta uma correlação positiva muito forte. Por outro lado, com os índices *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte.

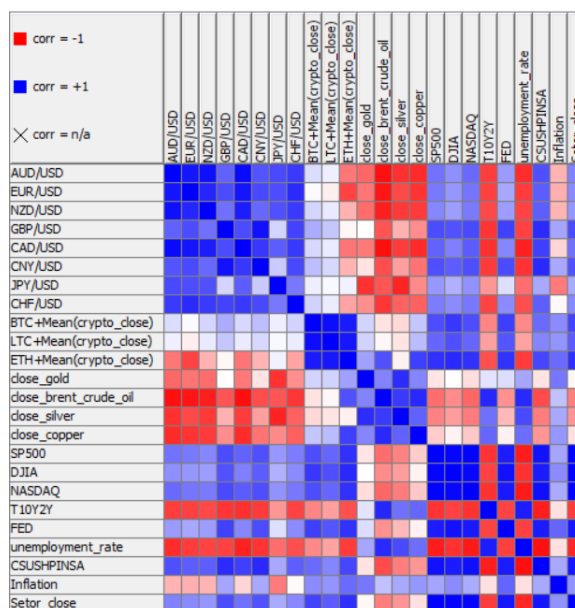


Figura 45 - Correlações com o setor *Industrials*

– *General Electric*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *General Electric* não apresenta qualquer correlação com o seu volume médio diário e com os pares *NZD/USD*, *GBP/USD* e *JPY/USD*. Por outro lado, verifica-se uma correlação positiva fraca entre o preço de fecho da *General Electric* com os pares *CNY/USD*, *CAD/USD*, *AUD/USD*, *EUR/USD* e *CHF/USD*.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *General Electric* tem uma correlação negativa moderada com a *LTC* e a *BTC*. Já com a *ETH* esta apresenta uma correlação negativa forte (-0,85).

Em termos de *commodities*, é possível notar que o preço de fecho da *General Electric* não apresenta qualquer tipo de correlação com a prata, sendo que o valor da correlação é inferior a 0,3 em valor absoluto. Já com o ouro e com o preço do barril de petróleo apresenta uma correlação negativa fraca e com o cobre uma correlação negativa moderada (-0,66).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *General Electric* não apresenta qualquer tipo de correlação com nenhum dos índices de mercado e com o índice do setor, apresentando valores absolutos inferiores a 0,3.

Por fim, é possível verificar que o preço de fecho da *General Electric* não apresenta qualquer tipo de correlação com nenhum dos fatores mencionados, apresentando valores absolutos inferiores a 0,3.

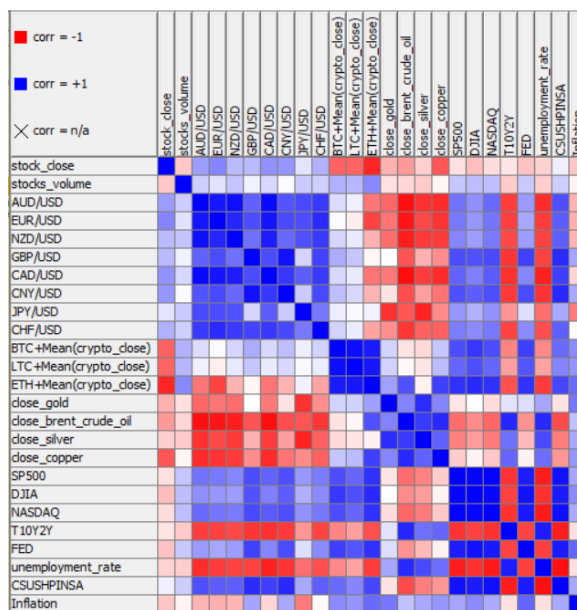


Figura 46 - Correlações com a empresa *General Electric*

9) Setor Information Technology

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com os pares *EUR/USD*, *NZD/USD*, *JPY/USD* e *CHF/USD*. Por outro lado, verifica-se uma correlação positiva moderada com os pares *AUD/USD*, *CAD/USD* e *CNY/USD* e uma correlação forte com o par *GBP/USD* (0,70).

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva moderada com a *LTC*. Já com a *ETH* e *BTC* esta apresenta uma correlação positiva forte.

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro e com o cobre. Já com a prata apresenta uma correlação negativa fraca e com o preço do barril de petróleo apresenta uma correlação negativa moderada.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva muito forte com todos os índices de mercado, apresentando valores entre 0,9 e 1.

Por fim, é possível verificar que o índice do setor apresenta uma correlação positiva fraca com a inflação. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte. Com o *FED* e com o índice *CSUSHPINSA* apresenta uma correlação positiva muito forte.

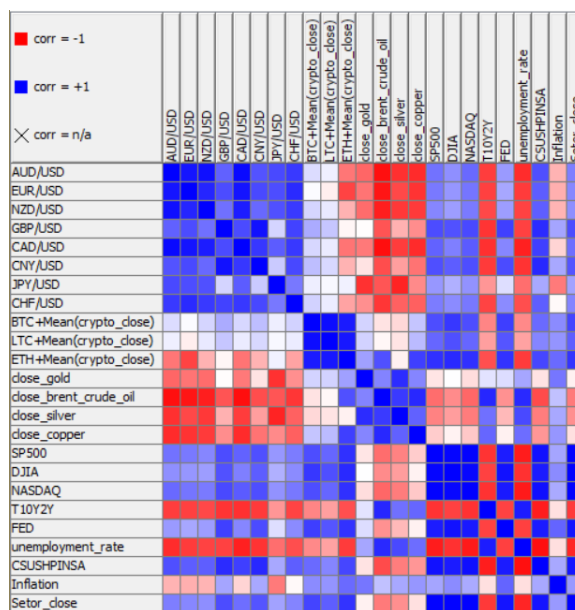


Figura 47 - Correlações com o setor *Information Technology*

– Apple

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Apple* apresenta uma qualquer correlação negativa moderada com o seu volume médio diário. Já com todos os pares de moedas estudados, o preço de fecho da *Apple* apresenta uma correlação positiva moderada, sendo todos os valores entre 0,5 e 0,7.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Apple* tem uma correlação positiva moderada com a *LTC* e a *BTC*. Já com a *ETH* esta apresenta uma correlação positiva forte (0,76).

Em termos de *commodities*, é possível notar que o preço de fecho da *Apple* não apresenta qualquer tipo de correlação com o ouro e com o cobre. Já com a prata e com o preço do barril de petróleo apresenta uma correlação negativa moderada.

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Apple* apresenta uma correlação positiva muito forte com todos os índices de mercado e com o índice do setor, apresentando valores entre 0,9 e 1.

Por fim, é possível verificar que o preço de fecho da *Apple* não apresenta qualquer tipo de correlação com a inflação, apresentando valores inferiores a 0,3. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte. Com o *FED* e com o índice *CSUSHPinsa* apresenta uma correlação positiva forte.

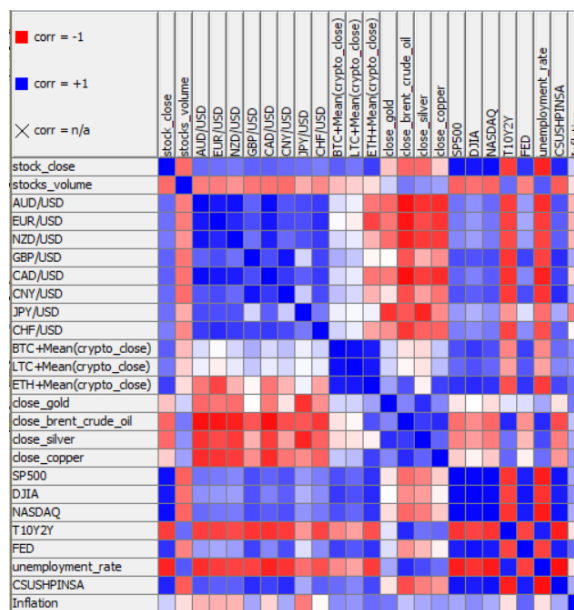


Figura 48 - Correlações com a empresa *Apple*

10) Setor *Materials*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com os pares *AUD/USD*, *EUR/USD*, *NZD/USD*, *JPY/USD* e *CHF/USD*. Por outro lado, verifica-se uma correlação positiva moderada com os pares *GBP/USD*, *CAD/USD* e *CNY/USD*.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva moderada com a *LTC*. Já com a *ETH* e *BTC* esta apresenta uma correlação positiva forte.

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro e com o cobre. Já com a prata e com o preço do barril de petróleo apresenta uma correlação negativa fraca.

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva muito forte com todos os índices de mercado estudados, observando-se valores superiores a 0,9.

Por fim, é possível verificar que o índice do setor apresenta uma correlação positiva fraca com a inflação. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte. Com o *FED* uma correlação positiva forte (0,88) e com o índice *CSUSHPINSA* uma correlação negativa muito forte (0,90).

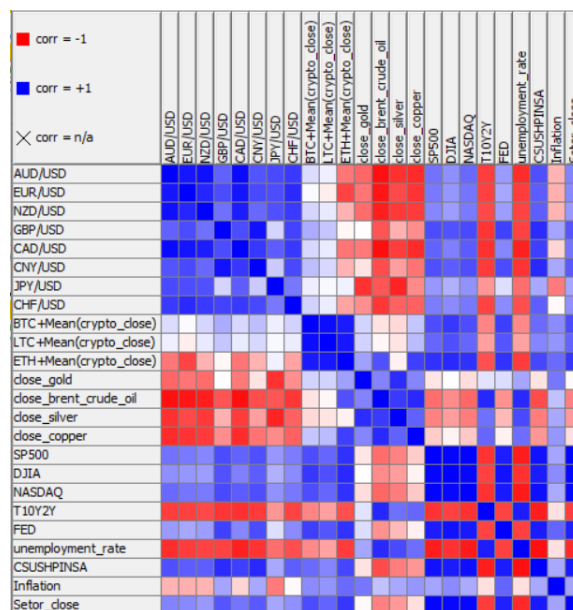


Figura 49 - Correlações com o setor *Materials*

– *Freeport-McMoRan*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Freeport-McMoRan* apresenta uma correlação negativa moderada com o seu volume médio diário e com o par *JPY/USD*. Com os pares *GBP/USD*, *CNY/USD* e *CHF/USD* apresenta uma correlação negativa forte. Já com os pares *AUD/USD*, *EUR/USD*, *NZD/USD* e *CAD/USD* apresenta uma correlação negativa muito forte, com valores superiores a 0,9.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Freeport-McMoRan* não tem qualquer tipo de correlação com a *LTC* e a *BTC*. Já com a *ETH* esta apresenta uma correlação positiva moderada (0,66).

Em termos de *commodities*, é possível notar que o preço de fecho da *Freeport-McMoRan* apresenta uma correlação positiva fraca com o ouro. Já com a prata e o cobre este apresenta uma correlação positiva forte e com o preço do barril de petróleo apresenta uma correlação positiva muito forte (0,96).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Freeport-McMoRan* apresenta uma correlação negativa fraca com os índices de mercado *DJIA* e com o índice do setor. Já com os índices de mercado *SP500* e *NASDAQ* apresenta uma correlação negativa moderada.

Por fim, é possível verificar que o preço de fecho da *Freeport-McMoRan* não apresenta qualquer tipo de correlação com a inflação (0,19). Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação positiva forte. Com o *FED* uma correlação negativa fraca e com o índice *CSUSHPINSA* uma correlação negativa forte (0,73).

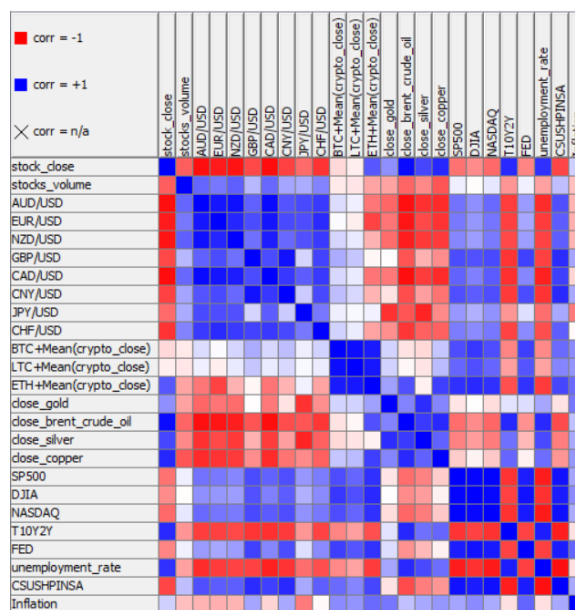


Figura 50 - Correlações com a empresa *Freeport-McMoRan*

11) Setor *Real Estate*

Relativamente aos pares de moedas, é possível constatar que o índice do setor apresenta uma correlação positiva fraca com o par *JPY/USD*. Já com o par *NZD/USD* apresenta uma correlação positiva moderada. Com todos os restantes apresenta uma correlação positiva forte.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva fraca com a *LTC*, a *BTC* e a *ETH*, apresentando valores entre 0,3 e 0,5.

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro. Já com a prata e com o cobre apresenta uma correlação negativa moderada e com o preço do barril de petróleo apresenta uma correlação negativa forte (-0,85).

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva forte com todos os índices de mercado estudados, com valores entre 0,7 e 0,9.

Por fim, é possível verificar que o índice do setor não apresenta qualquer tipo de correlação com a inflação. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa muito forte. Com o *FED* apresenta uma correlação positiva forte (0,73) e com o índice *CSUSHPINSA* uma correlação positiva muito forte (0,92).

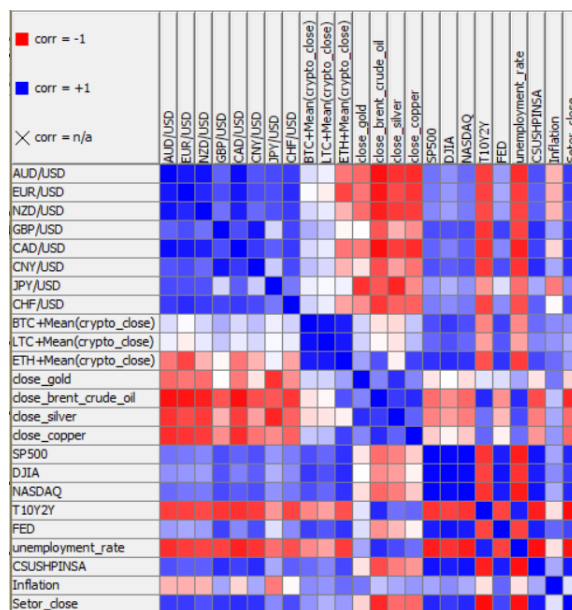


Figura 51 - Correlações com o setor *Real Estate*

– *Host Hotels & Resorts*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Host Hotels & Resorts* não apresenta qualquer correlação com o par *JPY/USD*. Já com o seu volume médio diário e com os pares *AUD/USD*, *EUR/USD*, *GBP/USD*, *CAD/USD*, *CNY/USD* e *CHF/USD* apresenta uma correlação negativa fraca. Com o par *NZD/USD* apresenta uma correlação negativa moderada (-0,53).

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Host Hotels & Resorts* não tem qualquer tipo de correlação com a *LTC* e a *BTC*. Já com a *ETH* esta apresenta uma correlação positiva moderada (0,63).

Em termos de *commodities*, é possível notar que o preço de fecho da *Host Hotels & Resorts* não apresenta qualquer tipo de correlação com o ouro e a prata. Já com o preço do barril de petróleo apresenta uma correlação positiva fraca e com o cobre uma correlação positiva moderada (0,56).

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Host Hotels & Resorts* não apresenta qualquer tipo de correlação com os vários índices de mercado e com o índice do setor.

Por fim, é possível verificar que o preço de fecho da *Host Hotels & Resorts* não apresenta qualquer tipo de correlação com a inflação, o índice *CSUSHPINSA*, a taxa de desemprego e as taxas de juros do FED. Já com o índice *T10Y2Y* apresenta uma correlação positiva fraca (0,31).

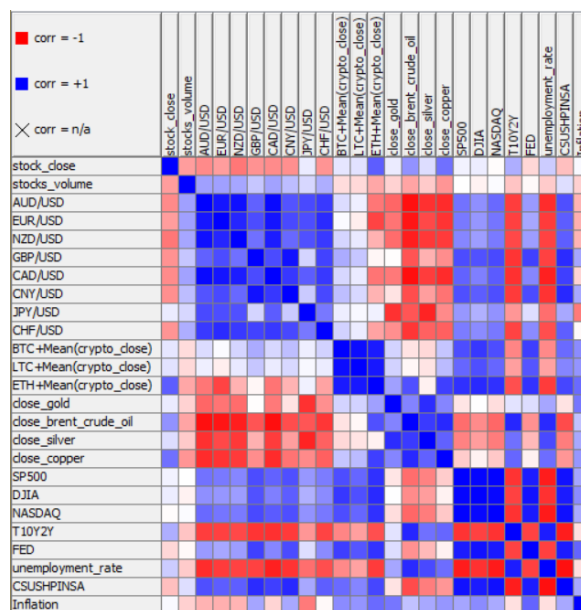


Figura 52 - Correlações com a empresa *Host Hotels & Resorts*

12) Setor *Utilities*

Relativamente aos pares de moedas, é possível constatar que o índice do setor não apresenta qualquer tipo de correlação com o par *JPY/USD*. Já com o par *NZD/USD* apresenta uma correlação positiva fraca. Com os pares *AUD/USD*, *EUR/USD*, *CAD/USD* e *CHF/USD* apresenta uma correlação positiva moderada. Por último, com os pares *GBP/USD* e *CNY/USD* apresenta uma correlação positiva forte.

Quanto às criptomoedas, é possível averiguar que o índice do setor tem uma correlação positiva fraca com a *LTC* (0,45). Já com a *BTC* e com a *ETH* este apresenta uma correlação positiva moderada.

Em termos de *commodities*, é possível notar que o índice do setor não apresenta qualquer tipo de correlação com o ouro. Já com a prata e o cobre apresenta uma correlação negativa fraca e com o preço do barril de petróleo uma correlação negativa moderada (-0,63).

No que diz respeito aos índices de mercado, é possível comprovar que o índice do setor apresenta uma correlação positiva forte com os índices de mercado *DJIA* e *NASDAQ*. Já com o índice de mercado *SP500* apresenta uma correlação positiva muito forte (0,90).

Por fim, é possível verificar que o índice do setor apresenta uma correlação positiva fraca com a inflação. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa forte. Com o *FED* apresenta uma correlação positiva forte (0,85) e com o índice *CSUSHPINSA* uma correlação positiva muito forte (0,91).

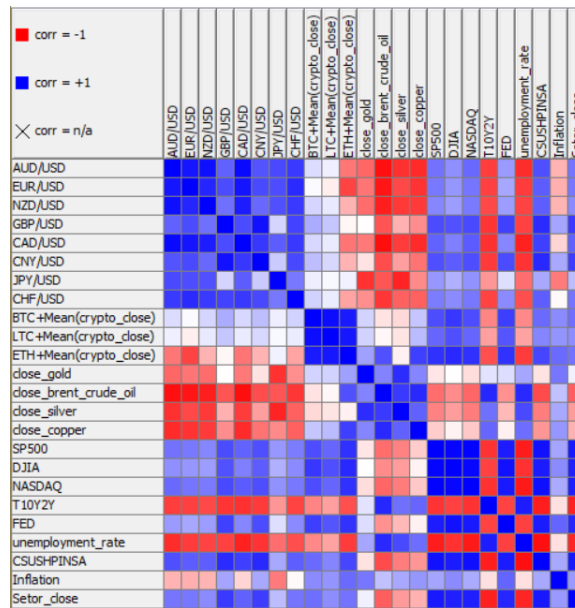


Figura 53 - Correlações com o setor *Utilities*

– *Exelon*

Relativamente aos pares de moedas, é possível constatar que o preço de fecho da *Exelon* não apresenta qualquer correlação com o seu volume médio diário e com os pares *AUD/USD*, *EUR/USD*, *NZD/USD*, *CAD/USD*, *JPY/USD* e *CHF/USD*. Já com os pares *GBP/USD* e *CNY/USD* apresenta uma correlação positiva fraca, tendo valores entre 0,3 e 0,5.

Quanto às criptomoedas, é possível averiguar que o preço de fecho da *Exelon* tem uma correlação positiva fraca com a *LTC* (0,43). Já com a *BTC* e com a *ETH* esta apresenta uma correlação positiva moderada.

Em termos de *commodities*, é possível notar que o preço de fecho da *Exelon* não apresenta qualquer tipo de correlação com nenhum dos fatores mencionados, tendo valores absolutos inferiores a 0,3.

No que diz respeito aos índices de mercado, é possível comprovar que o preço de fecho da *Exelon* apresenta uma correlação positiva moderada com os índices de mercado *SP500* e *NASDAQ*. Já com o índice de mercado *DJIA* e com o índice do setor apresenta uma correlação positiva forte, com valores entre 0,7 e 0,9.

Por fim, é possível verificar que a *Exelon* apresenta uma correlação positiva fraca com a inflação. Já com o índice *T10Y2Y* e com a taxa de desemprego apresenta uma correlação negativa moderada. Por último, com o *FED* e com o índice *CSUSHPINSA* uma correlação positiva moderada.

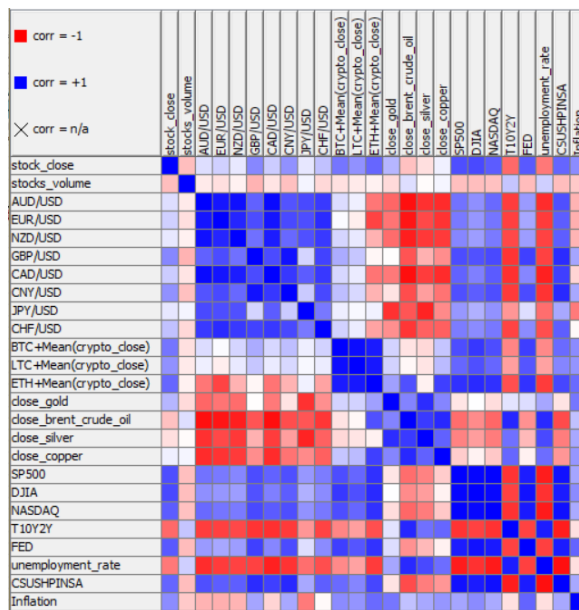


Figura 54 - Correlações com a empresa *Exelon*

4.2.4. Conclusão do Data Understanding

Após realizadas todas as correlações descritas, procedeu-se a uma análise dos resultados obtidos. Com objetivo de obter os melhores atributos, que permitissem posteriormente tecer melhores resultados na previsão, foi definido um parâmetro mínimo de correlação. Desta forma, para cada ação que se procurará prever, será incorporado todos os atributos em que a ação apresente uma correlação forte ou muito forte, assim como os que apresente uma correlação muito forte com o índice do setor a que pertence.

Através do *Tabela 13*, é possível constatar um resumo dos fatores que influenciaram a escolha das ações a que posteriormente se procedeu à sua previsão.

Tabela 13 - Quadro Resumo de Análise

EMPRESA	PERFORMANCE DA EMPRESA	PERFORMANCE DO SETOR	VOLATILIDADE	CLASSIFICAÇÃO (EMPRESA/SETOR) ^a
Facebook	567,83%	150%	Baixa	Bull / Bull
Ford Motor Company	-21,23%	71%	Baixa	Bear / Bull
Coca-Cola Company	5,49%	22%	Baixa	- / Bull
Kinder Morgan	-55,55%	0%	Alta	Bear / -
Bank of America	152,42%	88%	Média	Bull / Bull
Pfizer	14,95%	103%	Baixa	- / Bull
General Eletric	-31,42%	90%	Alta	Bear / Bull
Apple	159,64%	116%	Alta	Bull / Bull
Freeport-McMoRan	-40,27%	75%	Alta	Bear / Bull
Host Hotels & Resorts	8,96%	27%	Alta	- / Bull
Exelon	-4,31	21%	Média	- / Bull

^a “Bear” ou desempenho decrescente: $[-\infty, -20]\%$; “Bull” ou desempenho crescente: $[20, +\infty]\%$; “-” ou desempenho constante: $[-20, 20]\%$.

Após obter-se estes resultados, decidiu-se estudar duas empresas de cada classificação anteriormente dada, sendo que para cada classificação optou-se por escolher uma que apresente uma volatilidade alta e outra com volatilidade baixa. Para o caso da classificação (Empresa- “Bull” e Setor – “Bull”) optou-se pela escolha da *Facebook* e da *Apple*. Já para o caso da classificação (Empresa – “Bear” e Setor – “Bull”) escolheu-se a *Ford Motor Company* e a *Freeport McMoRan*. Por outro lado, para a classificação (Empresa – “Bear” e Setor – “-”) escolheu-se a *Kinder Morgan* apenas, pois não se verificou este comportamento em outras empresas. Por fim, para a classificação (Empresa – “-” e Setor – “Bull”) escolheu-se as empresas *Coca-Cola Company* e *Host Hotels & Resorts*.

Também se procederá à previsão do índice *SP500* de forma poder-se concluir qual dos ativos é mais fácil de prever, um índice ou uma ação individual. Para a construção do dataset de suporte à previsão do índice, será usado apenas variáveis que apresentam uma correlação forte e muito forte com o índice.

4.3. Data Preparation

Os dados recolhidos no início da fase *Data Understanding*, ainda apresentavam pequenos erros que precisavam de ser corrigidos para a construção do conjunto de dados final a ser utilizado na fase de *Modeling*. Através do *Apêndice I*, é possível visualizar toda a lógica realizada nesta fase, através do software KNIME.

O primeiro passo é efetivamente o tratamento dos valores em falta. Para isso foi utilizado o nodo “*Statistics*” que permitiu saber quais os atributos que apresentavam valores em falta, assim como a sua quantidade. Constatou-se a existência de valores em falta para todas as *commodities* e pares de moedas analisados, o índice *T10Y2Y* e por fim a criptomoeda *ETH*. Para todos eles, exceto a *ETH*, foram aplicados o valor fechado no dia anterior. Já para a *ETH*, como os valores em falta devem se ao facto de esta ainda estar numa fase embrionária e com um valor substancialmente reduzido, pelo que foi dado o valor fixo de zero.

De seguida, são descritas as variáveis que compõem o *dataset* de cada ação ou índice SP500, e que dessa forma alimentarão os modelos de previsão:

Tabela 14 - Variáveis para o *dataset* do *Data Understanding*

EMPRESA/ÍNDICE	PARES DE MOEDA	CRIPTOMOEDAS	ÍNDICES DE MERCADO	COMMODITIES	OUTRAS VARIÁVEIS
SP500	<i>GBP/USD</i>	<i>ETH</i>	<i>DJIA, NASDAQ</i>		<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Facebook	<i>GBP/USD, CAD/USD, CNY/USD</i>	<i>ETH</i>	<i>SP500, DJIA, NASDAQ</i>		<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Ford Motor Company	<i>GBP/USD, CAD/USD, CNY/USD</i>	<i>ETH</i>	<i>SP500, DJIA, NASDAQ</i>	barril de petróleo	<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Coca-Cola Company			<i>SP500, DJIA, NASDAQ</i>		<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Kinder Morgan	<i>GBP/USD, CAD/USD, CNY/USD</i>				<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Apple		<i>ETH</i>	<i>SP500, DJIA, NASDAQ</i>		<i>T10Y2Y</i> , FED, Taxa de desemprego, <i>CSUSHPINSA</i>
Freeport McMoRan	<i>AUD/USD, EUR/USD, NZD/USD, GBP/USD, CAD/USD, CNY/USD, CHF/USD</i>		<i>SP500, DJIA, NASDAQ</i>	barril de petróleo, prata e cobre	<i>T10Y2Y</i> , Taxa de desemprego, <i>CSUSHPINSA</i>
Host Hotels & Resorts					<i>T10Y2Y</i> , Taxa de desemprego, <i>CSUSHPINSA</i>

4.4. Modeling

Nesta fase, encontram-se identificadas os três modelos regressivos testados para prever o valor do índice *SP500* e das respectivas ações supra referidas para o dia seguinte: *Linear Regression*, *Random Forest* e o *MultiLayer Perceptron*. Estes modelos foram escolhidos tendo em consideração a análise realizada previamente na revisão bibliográfica e pelos modelos disponibilizados pela aplicação para problemas de regressão linear. Através do *Apêndice M*, é possível visualizar toda o *workflow* realizado nesta fase, através do software KNIME.

4.4.1. Construção dos Modelos

Com o auxílio do nó "*Partitioning*", os dados foram divididos em dois conjuntos: um conjunto de treino (70%) para a aprendizagem dos modelos e outro conjunto de teste (30%) que permitirá testar a sua aprendizagem. É ainda de realçar que o modo de seleção dos dados foi dado de forma aleatória, usando assim a opção *Draw Randomly* do nó mencionado.

Numa fase antecedente à previsão dos dados, torna-se essencial que o modelo aprenda a fazer essa previsão, a partir do conjunto de treino. Depois de definido o modelo a utilizar, os dados podem ser previstos com base no método escolhido. De seguida, os resultados do modelo são mostrados através do nó "*Numeric Scorer*", que avalia as previsões efetuadas.

Usou-se o node "*Window Slider*", de forma a obter os dados do dia anterior e dos três dias anteriores, com a finalidade de prever o dia seguinte, e isto, tanto para a variável target como para os restantes atributos incluídos na previsão. De seguida usou-se o nodo "*Joiner*", para englobar numa só tabela todos os dados obtidos com o nodo "*Window Slider*".

Em termos de configurações dos modelos, exclusivamente para a *Linear Regression* foram usadas as pré estabelecidas pelo KNIME por não se verificarem grandes alterações com o parâmetro possível de ser alterado.

Já no caso do *Random Forest Regression* e do *MultiLayer Perceptron* foi realizado uma análise de forma a perceber quais seriam os melhores parâmetros a usar tendo como função objetivo a minimização do *Root Mean Squared Error* (RMSE) resultante do *Numeric Scorer*, pois esta métrica pesa mais os grandes erros do que os pequenos erros e favorece um modelo que pode reagir a mudanças a curto prazo. Como se procederá duas experiências em termos de base temporal, uma que considera apenas o dia anterior à previsão e na outra experiência os três dias anteriores à previsão de forma a perceber aquele que permitirá obter melhores resultados, foi realizada essa análise para as duas bases temporais descritas.

Para o caso do *Random Forest Regression* foi estudado o parâmetro número de modelos que representa o número de árvores de decisão a serem aprendidas. Para a maioria dos conjuntos de dados, um valor entre 100 e 500 produz bons resultados, mas o número ótimo depende dos dados e deve, portanto, ser sujeito à devida análise dos parâmetros.

Para o *MultiLayer Perceptron* foi estudado os parâmetros número máximo de iterações a serem aprendidas, número de camadas escondidas e número de neurónios escondidos em cada camada.

Para o caso do SP500 obtiveram-se os números registados nas *Tabelas 15 e 16*. Para o *Random Forest Regression*, o número de modelos que permite obter um RMSE menor é, para o caso de um

dia, 100, e para o caso de três dias 500. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia 600 iterações no máximo, com 2 camadas e 10 neurónios escondidos e para o caso de três dias é 700 iterações no máximo, com 2 camadas e 15 neurónios escondidos.

Tabela 15 - Análise de Sensibilidade do *Random Forest* no *SP500*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	14.953	15.964
100	14.913	15.917
150	15.102	15.845
200	15.241	15.712
250	15.413	15.628
300	15.565	15.530
350	15.838	15.436
400	15.929	15.376
450	15.698	15.280
500	15.800	15.045

Tabela 16 - Análise de Sensibilidade do *MultiLayer Perceptron* no *SP500*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	22.225	14.711
550	2	10	21.857	15.051
600	2	10	21.690	15.304
650	2	10	22.093	14.795
700	2	10	22.303	14.664
500	3	10	21.779	14.867
550	3	10	22.284	15.141
600	3	10	22.190	15.348
650	3	10	22.082	14.897
700	3	10	22.370	15.396
500	2	15	21.808	15.457
550	2	15	22.361	14.800
600	2	15	21.966	14.983
650	2	15	21.819	15.002
700	2	15	22.159	14.635
500	3	15	21.732	14.677
550	3	15	22.415	15.579
600	3	15	21.980	15.125
650	3	15	22.206	14.759
700	3	15	21.886	14.824

Para o caso do *Facebook* obtiveram-se os números registados nas *Tabelas 17 e 18*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 500, e para o caso de três dias é o 450. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 650 iterações no máximo, com 2 camadas e 10 neurónios escondidos e para o caso de três dias é 550 iterações no máximo, com 3 camadas e 15 neurónios escondidos.

Tabela 17 - Análise de Sensibilidade do *Random Forest* na *Facebook*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	1.891	1.836
100	1.880	1.811
150	1.868	1.805
200	1.854	1.786
250	1.822	1.785
300	1.793	1.757
350	1.786	1.755
400	1.780	1.751
450	1.757	1.731
500	1.710	1.733

Tabela 18 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Facebook*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	2.404	2.157
550	2	10	2.321	2.255
600	2	10	2.622	2.217
650	2	10	1.970	2.195
700	2	10	2.410	2.237
500	3	10	2.323	2.428
550	3	10	2.177	2.324
600	3	10	2.312	2.196
650	3	10	1.979	2.124
700	3	10	2.206	2.139
500	2	15	2.182	2.182
550	2	15	2.997	2.276
600	2	15	3.075	2.101
650	2	15	2.920	2.098
700	2	15	2.901	2.102
500	3	15	2.091	2.227
550	3	15	2.391	2.097
600	3	15	2.277	2.253
650	3	15	2.599	2.099
700	3	15	2.649	2.585

Para o caso do *Ford Motor Company* obtiveram-se os números registados nas Tabelas 19 e 20. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 400, e para o caso de três dias é o 500. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 650 iterações no máximo, com 3 camadas e 10 neurónios escondidos e para o caso de três dias é 700 iterações no máximo, com 2 camadas e 15 neurónios escondidos.

Tabela 19 - Análise de Sensibilidade do *Random Forest* na *Ford Motor Company*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	0.223	0.217
100	0.221	0.217
150	0.221	0.216
200	0.220	0.217
250	0.219	0.216
300	0.212	0.210
350	0.204	0.208
400	0.202	0.207
450	0.212	0.207
500	0.214	0.206

Tabela 20 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Ford Motor Company*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	0.219	0.227
550	2	10	0.217	0.234
600	2	10	0.219	0.232
650	2	10	0.247	0.223
700	2	10	0.214	0.223
500	3	10	0.244	0.237
550	3	10	0.219	0.224
600	3	10	0.230	0.234
650	3	10	0.203	0.227
700	3	10	0.240	0.226
500	2	15	0.231	0.221
550	2	15	0.231	0.224
600	2	15	0.231	0.223
650	2	15	0.238	0.230
700	2	15	0.236	0.220
500	3	15	0.230	0.226
550	3	15	0.235	0.222
600	3	15	0.240	0.226
650	3	15	0.205	0.221
700	3	15	0.210	0.224

Para o caso do *Coca-Cola Company* obtiveram-se os números registados nas *Tabelas 21 e 22*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 300, e para o caso de três dias é o 450. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 600 iterações no máximo, com 2 camadas e 10 neurónios escondidos e para o caso de três dias é 550 iterações no máximo, com 3 camadas e 15 neurónios escondidos.

Tabela 21 - Análise de Sensibilidade do *Random Forest* na *Coca-Cola Company*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	0.435	0.467
100	0.429	0.453
150	0.428	0.453
200	0.425	0.451
250	0.424	0.450
300	0.422	0.449
350	0.423	0.449
400	0.424	0.477
450	0.425	0.446
500	0.424	0.477

Tabela 22 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Coca-Cola Company*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	0.427	0.468
550	2	10	0.426	0.451
600	2	10	0.422	0.465
650	2	10	0.427	0.488
700	2	10	0.453	0.484
500	3	10	0.458	0.434
550	3	10	0.436	0.458
600	3	10	0.434	0.455
650	3	10	0.434	0.441
700	3	10	0.428	0.437
500	2	15	0.433	0.442
550	2	15	0.438	0.468
600	2	15	0.444	0.449
650	2	15	0.446	0.483
700	2	15	0.433	0.462
500	3	15	0.438	0.430
550	3	15	0.437	0.428
600	3	15	0.444	0.438
650	3	15	0.434	0.444
700	3	15	0.433	0.441

Para o caso do *Kinder Morgan* obtiveram-se os números registados nas *Tabelas 23 e 24*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 150 e para o caso de três dias é o 450. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 650 iterações no máximo, com 3 camadas e 15 neurónios escondidos e para o caso de três dias é 650 iterações no máximo, com 3 camadas e 15 neurónios escondidos.

Tabela 23 - Análise de Sensibilidade do *Random Forest* na *Kinder Morgan*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	0.542	0.571
100	0.539	0.568
150	0.534	0.564
200	0.535	0.560
250	0.538	0.559
300	0.590	0.558
350	0.594	0.559
400	0.593	0.558
450	0.592	0.554
500	0.594	0.555

Tabela 24 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Kinder Morgan*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	0.556	0.659
550	2	10	0.549	0.630
600	2	10	0.605	0.637
650	2	10	0.577	0.628
700	2	10	0.624	0.622
500	3	10	0.560	0.628
550	3	10	0.570	0.623
600	3	10	0.553	0.637
650	3	10	0.603	0.622
700	3	10	0.561	0.630
500	2	15	0.559	0.624
550	2	15	0.638	0.640
600	2	15	0.627	0.651
650	2	15	0.629	0.628
700	2	15	0.607	0.628
500	3	15	0.564	0.624
550	3	15	0.603	0.624
600	3	15	0.620	0.630
650	3	15	0.547	0.621
700	3	15	0.610	0.630

Para o caso do *Apple* obtiveram-se os números registados nas *Tabelas 25 e 26*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 450, e para o caso de três dias é o 200. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia 600 iterações no máximo, com 3 camadas e 15 neurónios escondidos e para o caso de três dias é 700 iterações no máximo, com 2 camadas e 15 neurónios escondidos.

Tabela 5 - Análise de Sensibilidade do *Random Forest* na *Apple*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	1.592	1.731
100	1.593	1.741
150	1.594	1.728
200	1.595	1.704
250	1.596	1.706
300	1.593	1.714
350	1.592	1.719
400	1.592	1.717
450	1.588	1.724
500	1.592	1.725

Tabela 26 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Apple*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	1.803	6.012
550	2	10	1.831	5.958
600	2	10	2.013	5.841
650	2	10	1.987	5.935
700	2	10	1.900	5.831
500	3	10	1.850	5.792
550	3	10	1.842	5.846
600	3	10	1.851	5.874
650	3	10	1.771	5.819
700	3	10	1.829	5.988
500	2	15	1.942	5.967
550	2	15	1.792	5.826
600	2	15	1.800	5.933
650	2	15	1.845	5.927
700	2	15	1.903	5.778
500	3	15	1.809	5.947
550	3	15	1.838	5.878
600	3	15	1.757	5.981
650	3	15	1.854	5.923
700	3	15	2.177	5.892

Para o caso do *Freeport-McMoRan* obtiveram-se os números registados nas *Tabelas 27 e 28*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia, 100, e para o caso de três dias é o 400. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 600 iterações no máximo, com 2 camadas e 15 neurónios escondidos e para o caso de três dias é 550 iterações no máximo, com 3 camadas e 15 neurónios escondidos.

Tabela 6 - Análise de Sensibilidade do *Random Forest* na *Freeport-McMoRan*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	0.484	0.511
100	0.480	0.505
150	0.493	0.503
200	0.497	0.502
250	0.501	0.501
300	0.508	0.501
350	0.508	0.501
400	0.508	0.499
450	0.507	0.500
500	0.527	0.500

Tabela 287 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Freeport-McMoRan*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	0.667	0.572
550	2	10	0.593	0.628
600	2	10	0.529	0.612
650	2	10	0.610	0.572
700	2	10	0.549	0.574
500	3	10	0.527	0.625
550	3	10	0.642	0.622
600	3	10	0.594	0.602
650	3	10	0.543	0.587
700	3	10	0.594	0.573
500	2	15	0.535	0.586
550	2	15	0.567	0.582
600	2	15	0.524	0.690
650	2	15	0.531	0.572
700	2	15	0.528	0.580
500	3	15	0.526	0.578
550	3	15	0.535	0.569
600	3	15	0.538	0.572
650	3	15	0.539	0.580
700	3	15	0.782	0.578

Para o caso do *Host Hotels & Resorts* obtiveram-se os números registados nas *Tabelas 29 e 30*. Para o *Random Forest Regression*, o número de modelos que permite obter um *RMSE* menor é, para o caso de um dia é, 150, e para o caso de três dias é o 500. Já para *MultiLayer Perceptron*, o conjunto de parâmetros que permite obter um *RMSE* menor é, para o caso de um dia, 500 iterações no máximo, com 3 camadas e 15 neurónios escondidos e para o caso de três dias é 700 iterações no máximo, com 3 camadas e 10 neurónios escondidos.

Tabela 29 - Análise de Sensibilidade do *Random Forest* na *Host Hotels & Resorts*

NÚMERO DE MODELOS	RMSE (um dia)	RMSE (três dias)
50	0.299	0.297
100	0.296	0.292
150	0.294	0.292
200	0.295	0.291
250	0.299	0.290
300	0.302	0.287
350	0.302	0.287
400	0.303	0.286
450	0.303	0.286
500	0.303	0.283

Tabela 30 - Análise de Sensibilidade do *MultiLayer Perceptron* na *Host Hotels & Resorts*

NÚMERO MÁXIMO DE ITERAÇÕES	NÚMERO DE CAMADAS ESCONDIDAS	NÚMERO DE NEURÓNIOS ESCONDIDOS	RMSE (um dia)	RMSE (três dias)
500	2	10	0.293	0.295
550	2	10	0.289	0.299
600	2	10	0.298	0.298
650	2	10	0.298	0.301
700	2	10	0.295	0.304
500	3	10	0.291	0.302
550	3	10	0.289	0.297
600	3	10	0.286	0.292
650	3	10	0.300	0.293
700	3	10	0.292	0.291
500	2	15	0.285	0.292
550	2	15	0.289	0.293
600	2	15	0.287	0.294
650	2	15	0.293	0.292
700	2	15	0.288	0.296
500	3	15	0.279	0.292
550	3	15	0.285	0.299
600	3	15	0.295	0.294
650	3	15	0.287	0.296
700	3	15	0.286	0.293

Para todos os métodos suprarreferidos, foram considerados os seguintes alvos (“target”): preço de fecho do *SP500*, preço de fecho da empresa *Facebook*, *Ford Motor Company*, *Coca-Cola Company*, *Kinder Morgan*, *Apple*, *Freeport-McMoRan* e *Host Hotels & Resorts*.

4.5. Avaliação dos Modelos - Evaluation

Esta fase consiste em mostrar as métricas de avaliação dos modelos e comparar o seu desempenho com modelos semelhantes testados, e isto, com o objetivo de identificar qual o melhor modelo de previsão para cada uma das ações e índice *SP500*.

Tal como referido anteriormente usar-se-á as métricas do “*Numeric Scorer*” para a análise dos modelos de previsão utilizados. A principal métrica a ter em conta na análise no “*Numeric Scorer*” é o *RMSE* pelos motivos referenciados anteriormente. Contudo, também se analisará outras métricas para o melhor modelo encontrado, caso haja um, sendo essas métricas o R^2 por representar a variação da variável objetivo, em que 1 ou 100% é o melhor valor atingível, o *Mean Signed Difference (MSD)* por representar a média das diferenças entre a previsão e a variável objetivo, podendo ser negativo ou positivo dando assim uma ideia de direção do erro e por fim o *Mean Absolute Percentage Error (MAPE)*, por indicar o intervalo em que os valores de previsão se encontram face ao valor da variável objetivo.

Também foi criado um algoritmo de persistência como modelo de referência para realizar uma comparação dos resultados obtidos e proceder assim à devida conclusão relativamente às métricas obtidas. Este considera assim como previsão do preço do ativo para o dia seguinte, o preço do ativo no dia anterior à previsão, logo, se algum dos modelo de previsão estudado obtiver melhores resultados que o modelo de persistência, será considerado bom.

4.5.1. SP500

Tal como mostrado na *Tabela 31*, o modelo de *Linear Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever o índice *SP500* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 14.180, inferiores aos 14.330 do modelo de persistência e também um *MSD* mais perto de 0.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior tende a apresentar melhores resultados que a previsão baseada nos três dias anteriores. O único modelo em que isso não se verifica é o *MultiLayer Perceptron*.

Através do *Apêndice A*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 31 - Resultados do SP500

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	14.330	0.997	-0,581	0.005
Linear Regression (um dia)	14.180	0.997	0.392	0.005
Linear Regression (três dias)	14.555	0.997	0,597	0.005
Random Forest Regression (um dia)	14.913	0.997	0.685	0.005
Random Forest Regression (três dias)	15.045	0.997	0.544	0.005
MultiLayer Perceptron (um dia)	21.690	0.996	1.479	0.006
MultiLayer Perceptron (três dias)	14.555	0.997	0.597	0.005

4.5.2. Facebook

Tal como mostrado na *Tabela 32*, o modelo de *Linear Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever a ação da empresa *Facebook* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 1.645, inferiores aos 1.648 do modelo de persistência e também um *MSD* mais perto de 0.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior apresenta melhores resultados que a previsão baseada nos três dias anteriores.

Através do *Apêndice B*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 32 - Resultados da Facebook

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	1.648	0.998	-0.085	0.013
Linear Regression (um dia)	1.645	0.998	0.042	0.013
Linear Regression (três dias)	1.648	0.998	0.046	0.014
Random Forest Regression (um dia)	1.710	0.998	-0.062	0.014
Random Forest Regression (três dias)	1.731	0.998	0.187	0.015
MultiLayer Perceptron (um dia)	1.970	0.998	-0.047	0.019
MultiLayer Perceptron (três dias)	2.097	0.998	-0.054	0.020

4.5.3. Ford Motor Company

Tal como mostrado na *Tabela 33*, não existe um modelo estudado que apresente melhores resultados que o algoritmo de persistência para prever a ação da empresa *Ford Motor Company*. Contudo de todos os modelos estudados, aquele que apresentou melhores resultados foi o *Linear Regression* com base temporal de um dia.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior apresenta melhores resultados que a previsão baseada nos três dias anteriores.

Através do *Apêndice C*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 33 - Resultados da *Ford Motor Company*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	0.196	0.990	-0.012	0.010
Linear Regression (um dia)	0.199	0.990	-0,016	0.010
Linear Regression (três dias)	0.206	0.989	-0.019	0.011
Random Forest Regression (um dia)	0.202	0.989	-0.001	0.011
Random Forest Regression (três dias)	0.219	0.988	0.000	0.011
MultiLayer Perceptron (um dia)	0.203	0.989	-0.013	0.011
MultiLayer Perceptron (três dias)	0.220	0.987	-0.017	0.012

4.5.4. Coca-Cola Company

Tal como mostrado na *Tabela 34*, o modelo de *Linear Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever a ação da empresa *Coca-Cola Company* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 0.405, inferiores aos 0.408 do modelo de persistência, mas por outro lado um *MSD* mais longe de 0.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior tende a apresentar melhores resultados que a previsão baseada nos três dias anteriores. O único modelo em que isso não se verifica é o *MultiLayer Perceptron*.

Através do *Apêndice D*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 34 - Resultados da *Coca-Cola Company*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	0.408	0.971	0.023	0.007
Linear Regression (um dia)	0.405	0.971	0.038	0.007
Linear Regression (três dias)	0.408	0.971	0.042	0.007
Random Forest Regression (um dia)	0.422	0.969	0.073	0.007
Random Forest Regression (três dias)	0.446	0.965	0.076	0.007
MultiLayer Perceptron (um dia)	0.442	0.969	0.053	0.007
MultiLayer Perceptron (três dias)	0.428	0.968	0.054	0.007

4.5.5. Kinder Morgan

Tal como mostrado na *Tabela 35*, o modelo *Random Forest Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever a ação da empresa *Kinder Morgan* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 0.534, inferiores aos 0.539 do modelo de persistência, mas por outro lado um *MSD* mais longe de 0.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior apresenta melhores resultados que a previsão baseada nos três dias anteriores.

Através do *Apêndice E*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 35 - Resultados da *Kinder Morgan*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	0.539	0.996	0.004	0.014
Linear Regression (um dia)	0.540	0.996	-0.019	0.014
Linear Regression (três dias)	0.546	0.996	-0.027	0.015
Random Forest Regression (um dia)	0.534	0.997	0.018	0.015
Random Forest Regression (três dias)	0.546	0.996	-0.027	0.015
MultiLayer Perceptron (um dia)	0.547	0.996	-0.019	0.014
MultiLayer Perceptron (três dias)	0.621	0.995	0.016	0.017

4.5.6. Apple

Tal como mostrado na *Tabela 36*, o modelo *Random Forest Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever a ação da empresa *Apple* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 1.588, inferiores aos 1.660 do modelo de persistência, mas por outro lado um *MSD* mais longe de 0.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior apresenta melhores resultados que a previsão baseada nos três dias anteriores.

Através do *Apêndice F*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 36 - Resultados da *Apple*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	1.660	0.997	-0.026	0.011
Linear Regression (um dia)	1.671	0.997	0.072	0.011
Linear Regression (três dias)	1.697	0.996	0.088	0.011
Random Forest Regression (um dia)	1.588	0.997	0.099	0.011
Random Forest Regression (três dias)	1.704	0.996	0.032	0.012
MultiLayer Perceptron (um dia)	1.757	0.996	0.082	0.011
MultiLayer Perceptron (três dias)	5.778	0.958	-0.188	0.014

4.5.7. Freeport-McMoRan

Tal como mostrado na *Tabela 37*, os modelos *Random Forest Regression* com base temporal de um dia e o *Linear Regression* com base temporal de um dia apresentam melhores resultados que o algoritmo de persistência, sendo assim considerados os melhores modelos para prever a ação da empresa *Freeport-McMoRan* quando comparado com os restantes modelos estudados. Quanto ao *Linear Regression* com base temporal de um dia, verifica-se um *RMSE* de 0.481, inferiores aos 0.484 do modelo de persistência, mas por outro lado um *MSD* mais longe de 0. Já o *Random Forest Regression* com base temporal de um dia apresenta um *RMSE* de 0.480, inferior aos 0.484 do modelo de persistência e aos 0.481 do *Linear Regression* com base temporal de um dia, um *MSD* igual ao modelo de persistência, mas um *MAPE* superior a ambos.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior apresenta melhores resultados que a previsão baseada nos três dias anteriores.

Através do *Apêndice G*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 8 - Resultados da *Freeport-McMoRan*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	0.484	0.998	0.022	0.023
Linear Regression (um dia)	0.481	0.998	0.025	0.023
Linear Regression (três dias)	0.489	0.998	0.029	0.024
Random Forest Regression (um dia)	0.480	0.998	0.022	0.024
Random Forest Regression (três dias)	0.499	0.998	0.018	0.026
MultiLayer Perceptron (um dia)	0.524	0.997	0.013	0.025
MultiLayer Perceptron (três dias)	0.569	0.997	0.000	0.027

4.5.8. Host Hotels & Resorts

Tal como mostrado na *Tabela 38*, o modelo de *Linear Regression* com base temporal de um dia é o único que apresenta melhores resultados que o algoritmo de persistência, sendo assim considerado o melhor modelo para prever a ação da empresa *Host Hotels & Resorts* quando comparado com os restantes modelos estudados. Este modelo apresenta um *RMSE* de 0.270, inferiores aos 0.271 do modelo de persistência e um *R²* maior.

Em termos de bases temporais, os resultados obtidos concluem que a previsão baseada no dia anterior tende a apresentar melhores resultados que a previsão baseada nos três dias anteriores. O único modelo em que isso não se verifica é o *Random Forest Regression*.

Através do *Apêndice H*, é possível visualizar os resultados obtidos através de gráficos de comparação e diferença entre o valor real e valor de previsão, em KNIME.

Tabela 38 - Resultados da *Host Hotels & Resorts*

MODELO	RMSE	R ²	MSD	MAPE
Modelo de Persistência	0.271	0.986	0.001	0.010
Linear Regression (um dia)	0.270	0.987	0.001	0.010
Linear Regression (três dias)	0.272	0.986	-0.002	0.010
Random Forest Regression (um dia)	0.294	0.984	-0.001	0.012
Random Forest Regression (três dias)	0.283	0.985	0.005	0.011
MultiLayer Perceptron (um dia)	0.279	0.986	-0.005	0.011
MultiLayer Perceptron (três dias)	0.291	0.984	0.003	0.011

4.6. Discussão de resultados

De uma forma geral verifica-se que os resultados obtidos para o índice *SP500* face às restantes ações apresentam menores valores de *MAPE*, mas por outro lado apresentam maiores valores de *RMSE*. Isto deve-se ao facto de o preço do *SP500* ser superior ao preço das ações individuais, pois apesar de obterem-se variações diárias muito similares e na maioria inferiores a 1%, em valor

absoluto uma variação de 1% no *SP500* é sempre mais representativo do que o valor de uma ação individual. Contudo, o principal objetivo que se pretende alcançar é verificar que um dataset baseado em fatores económicos permite prever as ações do índice *SP500* e do próprio índice, pelo que de seguida se procederá à descrição dos modelos que apresentam melhores modelos que o modelo de persistência, que se destaca por ser um modelo simples e difícil de superar (Coimbra & Pedro, 2013).

Relativamente ao índice *SP500* verifica-se que apenas considerando o dia anterior à previsão com o modelo *Linear Regression* é que se obtém melhores resultados que o modelo de persistência, através do *MultiLayer Perceptron* com uma diferença de 0.15, a maior registada em todas as previsões realizadas, o que poderá corroborar a hipótese de que a previsão de índices permite obter melhores resultados que a previsão de ações individuais.

Relativamente ao par de ações empresa “*Bull*” (performance crescente) e setor “*Bull*” (performance crescente), Facebook e Apple, verificam-se comportamentos um pouco diferenciados. Em ambos os casos existem modelos capazes de superar o modelo de persistência, embora a diferença na Apple é substancialmente mais representativa que na Facebook. A diferença é que a Facebook tem como melhor modelo o *Linear Regression* e a Apple, o *Random Forest Regression*. Em termos de bases temporais, em ambas se verifica que o uso do dia anterior à previsão apresenta melhores resultados que os três dias anteriores.

Relativamente ao par de ações empresa “*Bear*” (performance decrescente) e setor “*Bull*” (performance crescente), *Ford Motor Company* e *Freeport-McMoRan*, verificam-se comportamentos muito diferenciados. Quanto à *Ford Motor Company*, não foi encontrado um modelo capaz de superar o modelo de persistência. Já quanto à *Freeport-McMoRan*, verifica-se que tanto considerando o modelo *Linear Regression* com base temporal de um dia, como usando o *Random Forest Regression* com base temporal de um dia, obtêm-se resultados melhores que o modelo de persistência sendo as suas diferenças pouco significativas.

Relativamente ao par de ações empresa “-” (performance constante) e setor “*Bull*” (performance crescente), *Coca-Cola Company* e *Host Hotels & Resorts*, verificam-se comportamentos mais similares. Em ambas o modelo capaz de superar o modelo de persistência é o *Linear Regression* que tem como base temporal um dia, com diferenças similares, mas não muito significativas.

Relativamente ao par de ações empresa “*Bear*” (performance decrescente) e setor “-” (performance constante), onde está incluído apenas a empresa *Kinder Morgan*, verifica-se que o único modelo capaz de superar o modelo de persistência é o *Random Forest Regression* com base temporal de um dia, com uma diferença um pouco significativa face aos restantes resultados.

Pelos dados obtidos há uma maior probabilidade de se obterem melhores resultados através do uso exclusivo de dados relativos ao dia anterior do que utilizando dados relativos aos três dias anteriores. Todos os modelos que superaram os modelos de persistência apresentam esta característica, independentemente do grupo de ação que se insere o ativo e da sua volatilidade intrínseca.

5. CONCLUSÃO

Neste capítulo tecem-se as devidas conclusões da dissertação, assim como as limitações evidenciadas ao longo da mesma. Com este capítulo pretende-se de uma forma clara e concisa resumir as conclusões obtidas durante todo este processo, e por outro lado, dar uma contribuição à comunidade científica, no sentido de expressar as dificuldades sentidas e os pontos que poderão ser melhorados para que os posteriores estudos tenham mais clareza quanto aos erros a evitar.

5.1. Conclusões finais

A presente dissertação teve como principal objetivo o estudo da previsão do mercado de ações, utilizando ações do índice *SP500* e o próprio índice. Este tem se revelado um desafio para a comunidade científica ao longo da sua existência, e nos últimos anos tem se alcançado resultados interessantes embora ainda inconsistentes, devido ao facto do mercado de ações apresentar um comportamento estocástico. Inicialmente, foi realizada uma descrição e análise detalhada de cada variável presente no *dataset*, assim como as possíveis correlações no sentido de se perceber quais seriam os atributos mais correlacionáveis com as ações que apresentaram mais volume em cada setor do *SP500* e o índice *SP500*. Nesta parte, com o objetivo de escolher as ações que se submeteria aos modelos de previsão, detetou-se que havia diferentes performances entre as empresas e o setor à qual a mesmas estão inseridas, optando-se assim por classificar as ações previamente analisadas em três tipos de classificação: “*Bull/Bull*”, “*Bear/Bull*”, “*-/Bull*” e “*Bear/-*”. No primeiro inserem-se a *Facebook*, a *Apple* e a *Bank of America*, sendo as duas primeiras as escolhidas pelo segundo critério, o de volatilidade, onde as primeiras mostraram ter um comportamento distinto, sendo assim mais interessante de comparar os resultados obtidos na previsão. No segundo conjunto inserem-se a *Ford Motor Company*, *Freeport-McMoRan* e a *General Electric*, sendo as duas primeiras as escolhidas pelo critério da volatilidade, onde as primeiras mostraram ter um comportamento distinto, e a *Freeport-McMoRan* apresentou uma volatilidade mais significativa que a *General Electric*. No terceiro conjunto inserem-se a *Coca-Cola Company*, *Host Hotels & Resorts*, *Pfizer* e a *Exelon*, sendo as duas primeiras as escolhidas pelo critério da volatilidade, onde as primeiras mostraram ter um comportamento distinto, e a *Coca-Cola Company* apresentou uma volatilidade menos significativa que a *Pfizer*.

Posteriormente submeteu-se o *dataset* com os atributos mais correlacionáveis dos vários ativos aos modelos de *machine learning* utilizados quer para uma base temporal de um dia, quer para três dias anteriores à previsão. A primeira evidencia sobre os resultados obtidos na previsão, é que a maior diferença registada do modelo criado face ao modelo de persistência, foi através do índice *SP500*, o que poderá indicar que a previsão de índices permite obter melhores resultados que a previsão de ações individuais. Relativamente ao par de ações empresa “*Bull*” (performance crescente) e setor “*Bull*” (performance crescente), obteve-se melhores resultados através da previsão da *Apple* do que com a *Facebook*, o que poderá indicar que fatores macroeconómicos com uma base temporal de um dia reagem bem perante empresas relativamente voláteis e com performance crescente alinhada com o setor. Relativamente ao par de ações empresa “*Bear*” (performance decrescente) e setor “*Bull*” (performance crescente), obteve-se melhores resultados na *Freeport-McMoRan* que na *Ford Motor Company*, o que poderá indicar que fatores macroeconómicos com uma base temporal de um dia reagem bem perante empresas

relativamente voláteis e com performance decrescente e desalinhado com o setor. Relativamente ao par de ações empresa “-” (performance constante) e setor “Bull” (performance crescente), *Coca-Cola Company* e na *Host Hotels & Resorts*, obteve-se resultados bastante similares pelo que não se consegue fazer uma distinção clara entre ambas, face à volatilidade superior que a *Host Hotels & Resorts* apresenta. Relativamente ao par de ações empresa “Bear” (performance decrescente) e setor “-”(performance constante), onde esta incluído apenas a empresa *Kinder Morgan*, obteve-se um resultado significativo quando comparado com as restantes ações analisadas. Como não há um ativo para comparar neste grupo de ações, podemos apenas concluir que mais uma vez fatores macroeconómicos reagem bem perante um ativo com volatilidade alta.

Posto isto, podemos concluir que de facto foram encontrados modelos capazes de prever o índice *SP500* ou mesmo ações de empresas presentes no índice *SP500*, mas porque os resultados não se estendem a todas as ações analisadas, essa hipótese não pode ser totalmente corroborada. Dos oito ativos estudados, em 87,5% deles obteve-se um modelo capaz de prever o preço de ativo melhor que o modelo de persistência. Relativamente às diferenças entre o melhor modelo de cada ativo e o modelo de persistência é possível aferir que as ações com menor volatilidade apresentam diferenças menos significativas. Enquanto as empresas mais voláteis tendem a apresentar diferenças relativamente mais significativas.

5.2. Limitações e investigação futura

Tal como se verifica nos resultados obtidos, a previsão do mercado de ações é realmente um tema complexo, que mesmo com a elevada capacidade computacional existente atualmente, ainda é preciso ter em consideração muitos fatores para lá da análise macroeconómica. A análise financeira das empresas e a análise técnica, apresentam uma componente importante neste processo e como tal também devem ser consideradas, com o objetivo de atingir melhores resultados.

Outro ponto que poderá ter influenciado o resultado das previsões positivamente, é o facto da base de dados encontrada ser relativo a um período de grande crescimento do mercado na sua globalidade, o que poderá afetar a consistência dos resultados face à existência de um período de recessão financeira. Embora uma grande quantidade de dados exige maior capacidade computacional, de forma a testar a veracidade de muitas hipóteses verificadas, seria importante compreender se as correlações dos dados se mantêm em diferentes períodos económicos. De forma a reverter esta situação, foram analisadas ações que tiveram um desempenho pior, que de certa forma, não se verificou um fator suficientemente significativo para quebrar a consistência dos resultados, mas, ainda assim, seria uma vantagem ter uma base de dados mais expressiva.

Embora o objetivo deste projeto tenha sido obter conhecimento de modelos capazes de prever o mercado de ações para as datas do *dataset* encontrado, o ideal seria obter um *dataset* que fosse continuamente alimentado ao longo do tempo, de forma a validar se os modelos e correlações encontradas são atuais e consistentes.

Relativamente ao software utilizado, ele é o mais indicado para quem não possui um conhecimento avançado ao nível da programação, mas por outro lado ainda há margem de melhoramento nomeadamente ao nível da personalização, agilização da lógica, mensagens de feedback perante erros encontrados, assim como o próprio processo de debug, e quanto à

capacidade computacional. Muitas das conclusões carecem de amostra maior, o que se tornou uma limitação face à elevada complexidade e exigência computacional que o software KNIME não oferece.

Em termos de horizonte temporal, seria viável o estudo das previsões para um espaço temporal de longo prazo, sendo que apenas foi abordado o curto prazo. Já em termos de modelos, a exploração intensiva de modelos ANN, ou mais especificamente os modelos RNNs é também uma opção viável, sendo que estes modelos apresentaram em muitos trabalhos da área, bons resultados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Akhilesh, G., Scott, S., & Schimitt, K. R. (2021). Dow Jones Industrial Average (DJIA) Definition. Retrieved December 29, 2021, from Investopedia website: <https://www.investopedia.com/terms/d/djia.asp>
- Allen, H., & Taylor, M. (1990). Charts, Noise and Fundamentals in the London Foreign Exchange Market. *Economic Journal*, 100(400), 49–59. Retrieved from <https://econpapers.repec.org/RePEc:ecj:econjl:v:100:y:1990:i:400:p:49-59>
- Ananthi, M., & Vijayakumar, K. (2020). Stock market analysis using candlestick regression and market trend prediction (CKRM). *Journal of Ambient Intelligence and Humanized Computing*, (0123456789). <https://doi.org/10.1007/s12652-020-01892-5>
- B. Mark Smith. (2003). *A History of the Global Stock Market: From Ancient Rome to Silicon Valley*. Retrieved from <https://books.google.com.br/books?hl=pt-BR&lr=&id=KJBjzXLEQggC&oi=fnd&pg=PP6&dq=stock+market+history&ots=MFBsmpVP6Q&sig=bzMXSYppyJoTFh58qN6HosQPR0g#v=onepage&q=stock+market+history&f=false>
- Berthold, M. R., Cebon, N., Dill, F., Di Fatta, G., Gabriel, T. R., Georg, F., ... Wiswedel, B. (2006). KNIME: The konstanz information miner. *4th International Industrial Simulation Conference 2006, ISC 2006*, 11(1), 58–61. <https://doi.org/10.1145/1656274.1656280>
- Breiman, L. (1994). *Bagging Predictors*. 20.
- Chen, J., Scott, G., & Munichello, K. (2020). Blue Chip. Retrieved December 29, 2021, from Investopedia website: <https://www.investopedia.com/terms/b/bluechip.asp#citation-2>
- Clinton, J. (2000). *Step-by-step data mining guide*.
- Coimbra, C. F. M., & Pedro, H. T. C. (2013). Stochastic-Learning Methods. In *Solar Energy Forecasting and Resource Assessment*. <https://doi.org/10.1016/B978-0-12-397177-7.00015-2>
- Colin, S. (2000). The Crisp Dm Model the New Blueprint for Data Mining. Retrieved January 2, 2022, from <https://pdfslide.net/documents/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.html>
- Demšar, J., Erjavec, A., Hočevar, T., Milutinovič, M., Možina, M., Toplak, M., ... Zupan, B. (2013). Orange: Data Mining Toolbox in Python Tomaž Curk Matija Polajnar Laň Zagar. *Journal of Machine Learning Research*, 14, 2349–2353.
- Devi, K. N., & Bhaskaran, V. M. (2015). Semantic Enhanced Social Media Sentiments for Stock Market Prediction. *International Journal of Economics and Management Engineering*, 9(2), 684–688. Retrieved from <https://waset.org/publications/10001130/semantic-enhanced-social-media-sentiments-for-stock-market-prediction>
- Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. Retrieved January 10, 2022, from PESIT South Campus website: https://www.researchgate.net/publication/309492895_Forecasting_to_Classification_Predicting_the_direction_of_stock_market_price_using_Xtreme_Gradient_Boosting?channel=doi&linkId=581384c008aedc7d8961e371&showFulltext=true
- Di Persio, L., & Honchar, O. (2017). Recurrent neural networks approach to the financial forecast of Google assets. *INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTERS IN SIMULATION*, 11, 7.
- Dobson, J., & Soenen, L. (1993). Three Agency-Cost Reasons for Hedging Foreign Exchange Risk.

- Managerial Finance*, 19(6), 35–44. <https://doi.org/10.1108/EB013728>
- Dow Jones Industrial Average® | S&P Dow Jones Indices. (2021). Retrieved December 29, 2021, from spglobal website: <https://www.spglobal.com/spdji/en/indices/equity/dow-jones-industrial-average/#overview>
- Elleuch, J. (2009). Fundamental Analysis Strategy and the Prediction of Stock Returns. *International Research Journal of Finance and Economics*.
- Enke, D., Grauer, M., & Mehdiyev, N. (2011). Stock Market Prediction with Multiple Regression, Fuzzy Type-2 Clustering and Neural Networks. Retrieved January 14, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/S1877050911005035?token=000839A4C83EFC7219BF8D7356E983FA2DCEBF310E160AAC7B738F29EAECBBC36BE31CA731F14B0AF22BCABA721C6D20&originRegion=eu-west-1&originCreation=20220114221504>
- Falinouss, P. (2007). *MASTER'S THESIS Stock Trend Prediction Using News Articles A Text Mining Approach*.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383. <https://doi.org/10.2307/2325486>
- Fama, E. F. (1995). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 51(1), 75–80. <https://doi.org/10.2469/faj.v51.n1.1861>
- Fama, E., & French, K. R. (1993). Common Risk Factors in Returns on Stocks and Bonds. *Journal of Financial Economics*, 3–56.
- Farhat, N. H. (1992). Photonit neural networks and learning machines the role of electron-trapping materials. *IEEE Expert-Intelligent Systems and Their Applications*, 7(5), 63–72. <https://doi.org/10.1109/64.163674>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–37. <https://doi.org/10.1609/AIMAG.V17I3.1230>
- Ferreira, C. (2021a). *CRISP-DM: A Data Mining & Knowledge Discovery Process Model*.
- Ferreira, C. (2021b). *Data Mining and Knowledge Discovery*.
- Ferreira, C. (2021c). Predictive Models. <https://doi.org/10.1002/9781119209515.ch12>
- Floratos, A., Smith, K., Ji, Z., Watkinson, J., & Califano, A. (2010). Genome analysis geWorkbench: an open source platform for integrative genomics. *BIOINFORMATICS APPLICATIONS NOTE*, 26(14), 1779–1780. <https://doi.org/10.1093/bioinformatics/btq282>
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine*, 13(3), 57–57. <https://doi.org/10.1609/AIMAG.V13I3.1011>
- Ganser, A., Hollaus, B., & Stabinger, S. (2021). Classification of Tennis Shots with a Neural Network Approach. *Sensors 2021, Vol. 21, Page 5703*, 21(17), 5703. <https://doi.org/10.3390/S21175703>
- Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., ... Vincent, K. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8). <https://doi.org/10.1186/gb-2010-11-8-r86>
- Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. Retrieved January 12, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/O95418109400011S?token=CF49C61C52A9A3F7965F2AFB2E1E1BD49DB215242C6E2990DD69E0F0057C7DA8467C227E398D3F37C2F645245>

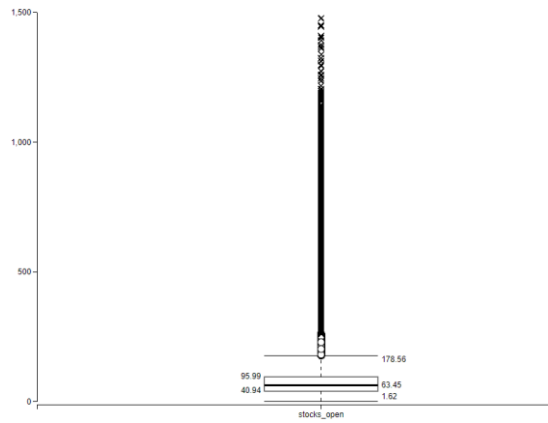
- DCE580C&originRegion=eu-west-1&originCreation=20220112221337
- Gupta, R., Garg, N., & Singh, S. (2013). Stock market prediction accuracy analysis using kappa measure. *Proceedings - 2013 International Conference on Communication Systems and Network Technologies, CSNT 2013*, 5. <https://doi.org/10.1109/CSNT.2013.136>
- Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction . Retrieved January 10, 2022, from Elsevier Enhanced Reader website:
<https://reader.elsevier.com/reader/sd/pii/S0957417411002740?token=726C66BA7C6A2462A22F530CCD13156C63CBC1483D5C8ABEA02DC33C1CB597318C69A7090033A0B06248175C8B28C172&originRegion=eu-west-1&originCreation=20220110232043>
- Gururaj, V., & R, S. V. (2019). Stock Market Prediction using Linear Regression and Support Vector Machines. *International Journal of Applied Engineering Research*, 14(8), 1931–1934. Retrieved from <http://www.ripublication.com>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Hellström, T., & Holmström, K. (1998). Predictable Patterns in Stock Returns. *UNIVERSITY, DEPARTMENT OF*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.3351>
- Horne, J. C., & Wachowicz, J. M. J. (2004). *Fundamentals of Financial Management* .
- Huang, C.-F., & Li, H.-C. (2017). *An Evolutionary Method for Financial Forecasting in Microscopic High-Speed Trading Environment*. <https://doi.org/10.1155/2017/9580815>
- Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). Taverna: a tool for building and running workflows of services. *Nucleic Acids Research*, 34(W729–W732). <https://doi.org/10.1093/nar/gkl320>
- Infopédia. (2021). Bolsa de Valores - Infopédia. Retrieved December 29, 2021, from Porto Editora website: [https://www.infopedia.pt/\\$bolsa-de-valores](https://www.infopedia.pt/$bolsa-de-valores)
- Jagla, B., Wiswedel, B., & Coppée, J.-Y. (2011). Extending KNIME for next-generation sequencing data analysis. *BIOINFORMATICS APPLICATIONS NOTE*, 27, 2907–2909. <https://doi.org/10.1093/bioinformatics/btr478>
- Jeffrey A. Franke; Kenneth A. Froot. (1990). *Chartists, Fundamentalists, and Trading in the Foreign Exchange Market*. 185. Retrieved from https://www.researchgate.net/publication/4721069_Chartists_Fundamentalists_and_Trading_in_the_Foreign_Exchange_Market
- Kenton, W., Boyle, M. J., & Munichiello, K. (2021). S&P 500 Index: Standard & Poor’s 500 Index Definition. Retrieved December 29, 2021, from Investopedia website: <https://www.investopedia.com/terms/s/sp500.asp#citation-21>
- Knime. (2019). KNIME Workbench Guide. Retrieved January 15, 2022, from KNIME Analytics Platform 4.0 website: https://docs.knime.com/2019-06/analytics_platform_workbench_guide/index.html#workspaces
- KNIME. (2021a). KNIME Community Forum. Retrieved December 30, 2021, from <https://forum.knime.com/>
- KNIME. (2021b). KNIME Learning . Retrieved December 30, 2021, from KNIME Learning website: <https://www.knime.com/learning>
- KNIME. (2021c). KNIME Quickstart Guide. Retrieved December 30, 2021, from KNIME

- Documentation website: https://docs.knime.com/2020-07/analytics_platform_quickstart_guide/index.html#introduction
- Kroha, P., & Baeza-Yates, R. (2004). *Classification of Stock Exchange News*.
- Linke, B., Giegerich, R., Goesmann, A., & Bateman, A. (2011). *Conveyor: a workflow engine for bioinformatic analyses*. 27(7), 903–911. <https://doi.org/10.1093/bioinformatics/btr040>
- Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., ... Zhao, Y. (2006). Scientific workflow management and the Kepler system. In *Concurrency and Computation: Practice and Experience* (Vol. 18). <https://doi.org/10.1002/CPE.994>
- Lugmayr, A., & Gossen, G. (2012). Evaluation of Methods and Techniques for Language Based Sentiment Analysis for DAX30 Stock Exchange - A First Concept of a “LUGO” Sentiment Indicator. Retrieved January 10, 2022, from Tampere University of Technology (TUT) website: https://www.researchgate.net/publication/259287571_Evaluation_of_Methods_and_Techniques_for_Language_Based_Sentiment_Analysis_for_DAX30_Stock_Exchange_-_A_First_Concept_of_a_LUGO_Sentiment_Indicator
- MacroTrends. (2021a). Dow Jones - DJIA - 100 Year Historical Chart . Retrieved January 15, 2022, from MacroTrends website: <https://www.macrotrends.net/1319/dow-jones-100-year-historical-chart>
- MacroTrends. (2021b). NASDAQ Composite - 45 Year Historical Chart. Retrieved January 15, 2022, from MacroTrends website: <https://www.macrotrends.net/1320/nasdaq-historical-chart>
- MacroTrends. (2021c). S&P 500 Index - 90 Year Historical Chart . Retrieved January 15, 2022, from MacroTrends website: <https://www.macrotrends.net/2324/sp-500-historical-chart-data>
- Mahmoud, A., and Sakr, S. (2012). The Predictive Power of Fundamental Analysis in Terms of Stock Return and Future Profitability performance. *International Research Journal of Finance and Economics*.
- Malkiel, B. G. (1996). *A random walk down Wall Street : including a life-cycle guide to personal investing*. Norton.
- Milosevic, N. (2016). Equity forecast: Predicting long term stock price movement using machine learning. *School of Computer Science*.
- More, A. M., Rathod, P. U., Patil, R. H., Sarode, D. R., & Student, B. E. (2018). Stock Market Prediction System using Hadoop. *International Journal of Engineering Science and Computing*. Retrieved from <http://ijesc.org/>
- Moro, S., Laureano, R. M. S., Cortez, P., & Algoritmi, C. (n.d.). *USING DATA MINING FOR BANK DIRECT MARKETING: AN APPLICATION OF THE CRISP-DM METHODOLOGY*.
- Muhammad, S. (2018). The Relationship Between Fundamental Analysis and Stock Returns Based on the Panel Data Analysis; Evidence from Karachi Stock exchange (KSE). *Research Journal of Finance and Accounting Wwww.iiste.Org ISSN, 9(3)*. Retrieved from www.iiste.org
- Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, 2(5–6), 183–197. [https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/10.1016/0925-2312(91)90023-5)
- Nayak, A., Pai, M. M. M., & Pai, R. M. (2016). Prediction Models for Indian Stock Market . Retrieved January 10, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/S1877050916311619?token=16DD3428FCB5466A0CC32BBCE6B3FEA00ECD5117883CD6E906F2E00062577541092663CC13D420F276470C1765E99FF8&originRegion=eu-west-1&originCreation=20220110233230>

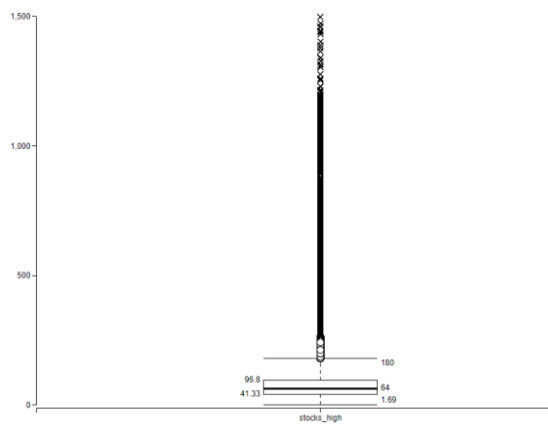
- Neely, C. J., White, G. L., & Street Journal, W. (1992). *Technical Analysis in the Foreign Exchange Market: A Layman's Guide*.
- Néron, B., Ménager, H., Maufrais, C., Joly, N., Maupetit, J., Letort, S., ... Letondal, C. (2009). *Mobyle: a new full web bioinformatics framework*. *25(22)*, 3005–3011. <https://doi.org/10.1093/bioinformatics/btp493>
- Northover, P. (1999). Evolutionary growth theory and forms of realism. *Cambridge Journal of Economics*, *23(1)*, 33–63. <https://doi.org/10.1093/CJE/23.1.33>
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7(1)*, 1–40. <https://doi.org/10.1186/S40537-020-00299-5/TABLES/25>
- Peachavanish, R. (2016). Stock Selection and Trading Based on Cluster Analysis of Trend and Momentum Indicators. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, *1*, 5.
- Piotroski J. (2002). Value investing: The use of Historical financial statement information to separate winners from losers. *Journal of Accounting Research*.
- Pring, M. J. (1991). *Technical Analysis Explained : The Successful Investor's Guide to Spotting Investment Trends and Turning Points*. Retrieved from https://books.google.com/books/about/Technical_Analysis_Explained.html?id=2mUPAQAA MAAJ
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). *The Effects of Twitter Sentiment on Stock Price Returns*. <https://doi.org/10.1371/journal.pone.0138441>
- Rankia Portugal. (2021). Maiores quedas históricas na bolsa . Retrieved December 29, 2021, from Rankia Portugal website: <https://www.rankia.pt/bolsa/maiores-quedas-historicas-na-bolsa/>
- Rizky Sabilla, A., & Kurniasih, A. (n.d.). *THE EFFECT OF MACROECONOMICS ON STOCK INDEX*.
- Robles, P. D., Ivan V, G., & Belik, I. V. (2010). A Comparison of Machine Learning Classifiers Applied to Financial Datasets. *Proceedings of the World Congress on Engineering and Computer Science*, *1*, 6.
- Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics (Switzerland)*, *10(21)*. <https://doi.org/10.3390/electronics10212717>
- Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing and Management*, *45(5)*, 571–583. <https://doi.org/10.1016/j.ipm.2009.05.001>
- Segal, T. (2021). Buttonwood Agreement Definition. Retrieved December 29, 2021, from <https://www.investopedia.com/terms/b/buttonwoodagreement.asp>
- Seng, D., & Hancock, J. R. (2012). Fundamental Analysis and the Prediction of Earnings. *International Journal of Business and Management*, *7(3)*. <https://doi.org/10.5539/ijbm.v7n3p32>
- Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies* , *7(2)*, 26. <https://doi.org/10.3390/IJFS7020026>
- Sharma, S., Ahmed, S., Naseem, M., Alnumay, W. S., Singh, S., & Cho, G. H. (2021). A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength

- Estimation in Construction and Geotechnical Engineering. *Sensors 2021, Vol. 21, Page 463, 21(2)*, 463. <https://doi.org/10.3390/S21020463>
- Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3), 275–294. <https://doi.org/10.1002/wics.1198>
- Tan, T. Z., Quek, C., & Ng, G. S. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23(2), 236–261. <https://doi.org/10.1111/j.1467-8640.2007.00303.x>
- The Investopedia team, & Anderson, S. (2021). The Dow vs. the Nasdaq: Key Differences to Know. Retrieved December 29, 2021, from Investopedia website: <https://www.investopedia.com/ask/answers/difference-between-dow-and-nasdaq/>
- Tretina, K., & Schmidt, J. (2021). What Is The Stock Market? – Forbes Advisor. Retrieved December 29, 2021, from <https://www.forbes.com/advisor/investing/what-is-the-stock-market/>
- Weng, B., Ahmed, M., & Megahed, F. (2017). Stock market one-day ahead movement prediction using disparate data sources. Retrieved January 10, 2022, from ELSEVIER website: <https://reader.elsevier.com/reader/sd/pii/S0957417417301331?token=65A6DC5BC0329471F41ECD24F85D0BC0DAD0CCCF3A3EFD022F180BEB8CFEBC4AC5DB940C1FBBBE33E18D0D741F853B35&originRegion=eu-west-1&originCreation=20220110223752>
- Whalen, S., Prakash Pandey, O., & Pandey, G. (n.d.). *Predicting protein function and other biomedical characteristics with heterogeneous ensembles HHS Public Access*. <https://doi.org/10.1016/j.ymeth.2015.08.016>
- Wirth, R., & Hipp, J. (n.d.). *CRISP-DM: Towards a Standard Process Model for Data Mining*.
- World Federation of Exchanges. (2020). Market capitalization of listed domestic companies (current US\$) | Data. Retrieved December 29, 2021, from <https://data.worldbank.org/indicator/CM.MKT.LCAP.CD>
- Xianya, J., Mo, H., & Haifeng, L. (2019). Stock Classification Prediction Based on Spark. Retrieved January 12, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/S1877050919319933?token=B8C919B484D82E97943500989343378B6AA6ED65664B00219F19D27B16FC4B7670EB62A4B8E3B1889EC8CB947FDEAF9C&originRegion=eu-west-1&originCreation=20220112230944>
- Yudong, Z., & Lenan, W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network . Retrieved January 10, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/S095741740800852X?token=C6D775283AFDCF8CC13DD39AFF959D791F77596DEF7FB3048187A02F4366CF5C7A3B3E9ED85BC7F943B2C0980C96EA10&originRegion=eu-west-1&originCreation=20220110231623>
- Zhang, F., Li, J., Wang, Y., Guo, L., Wu, D., Wu, H., ... Avizzano, A. (2021). *Ensemble Learning Based on Policy Optimization Neural Networks for Capability Assessment*. <https://doi.org/10.3390/s21175802>
- Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction . Retrieved January 12, 2022, from Elsevier Enhanced Reader website: <https://reader.elsevier.com/reader/sd/pii/S0957417416305115?token=DA40997B9192FE0A1A24A74D67EFAC48F9D0149A1769979F0978C1EAA30AD7FF73CAB2C536E85231502869FB2738A749&originRegion=eu-west-1&originCreation=20220112220941>

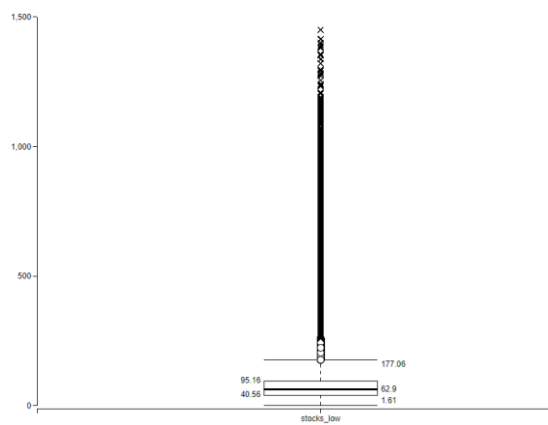
APÊNDICE A – BOXPLOTS DOS ATRIBUTOS DO DATASET INICIAL



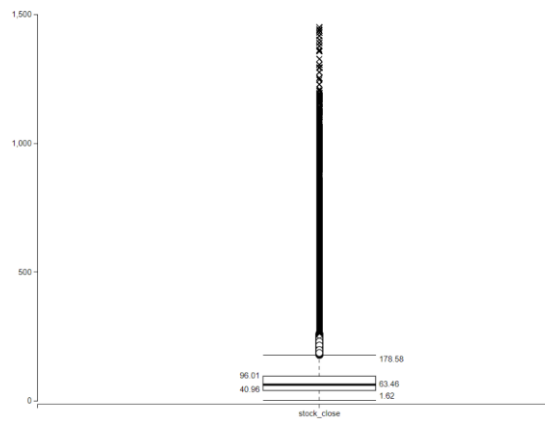
Boxplot da variação do preço de abertura



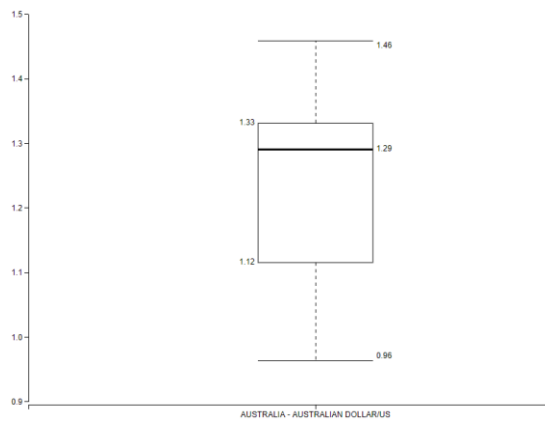
Boxplot da variação do preço mais alto



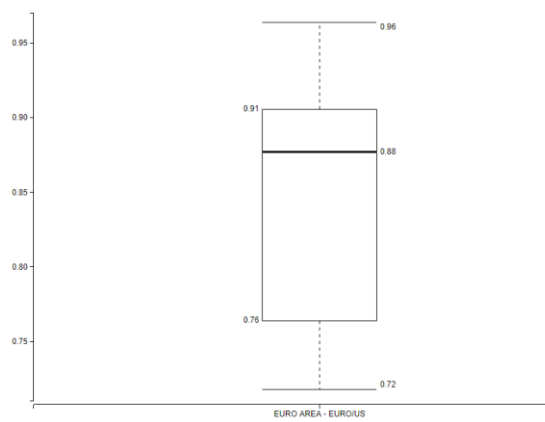
Boxplot da variação do preço mais baixo



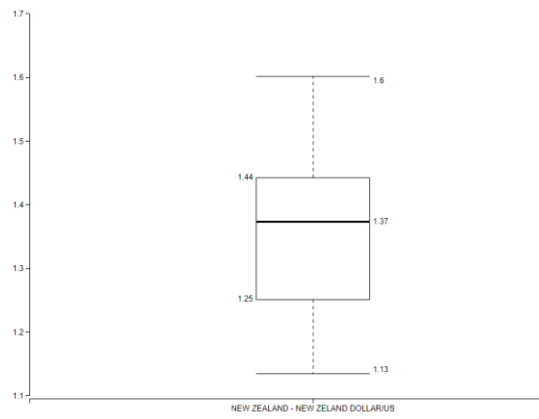
Boxplot da variação do preço de fecho



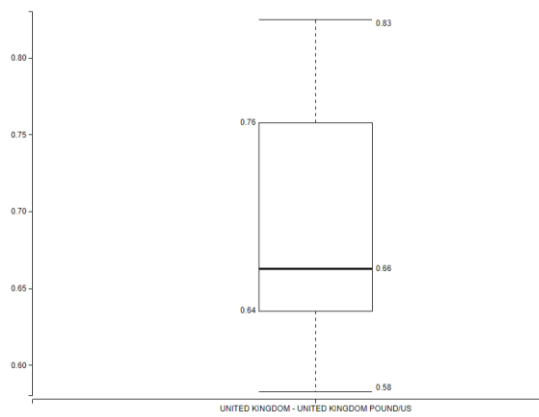
Boxplot da variação do preço do par *AUD/USD*



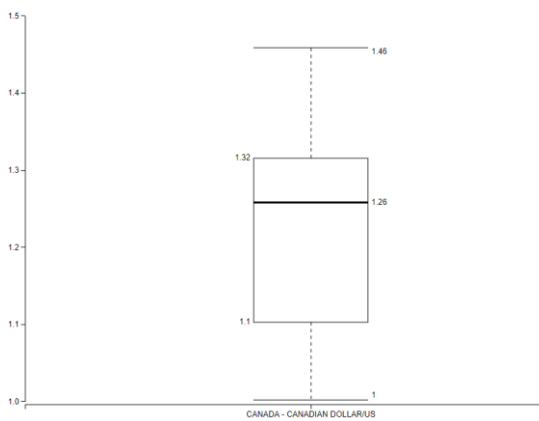
Boxplot da variação do preço do par *EUR/USD*



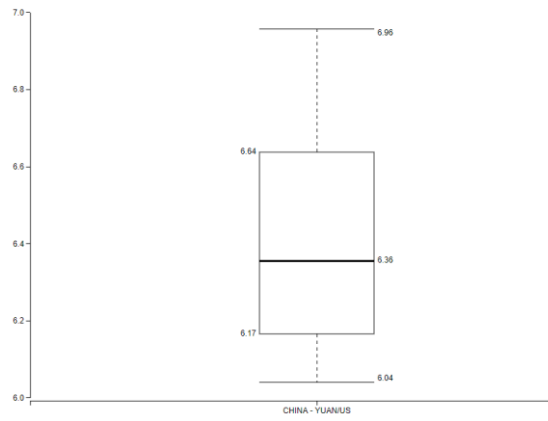
Boxplot da variação do preço do par *NZD/USD*



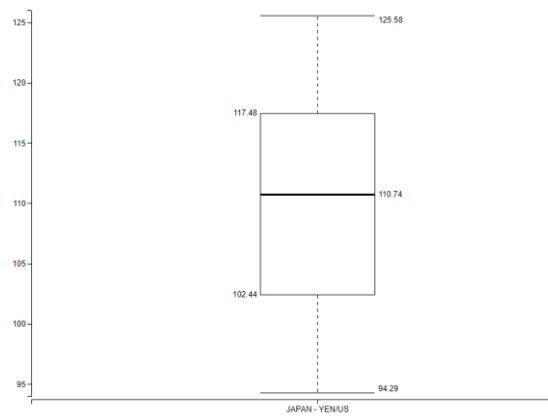
Boxplot da variação do preço do par *GBP/USD*



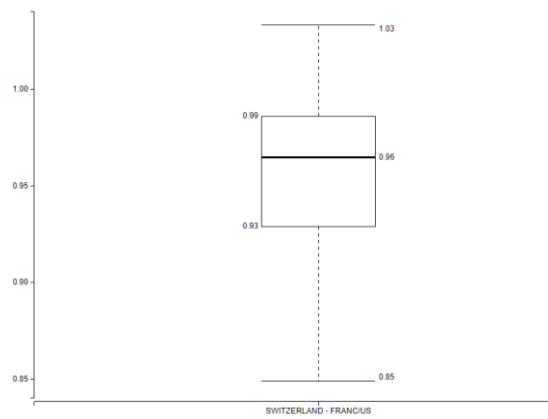
Boxplot da variação do preço do par *CAD/USD*



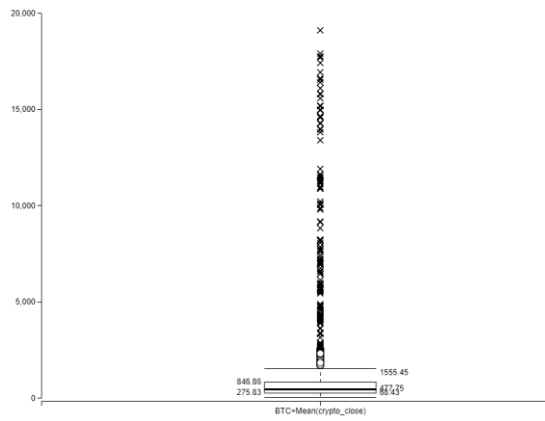
Boxplot da variação do preço do par *CNY/USD*



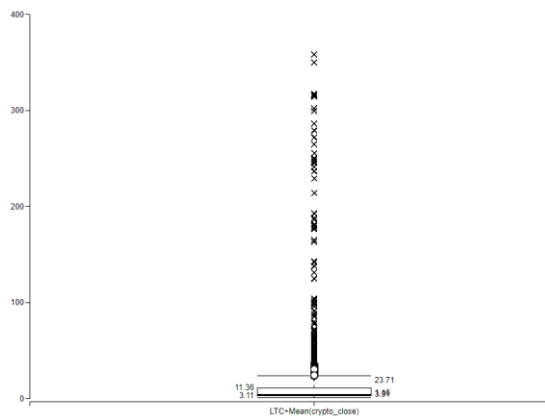
Boxplot da variação do preço do par *JPY/USD*



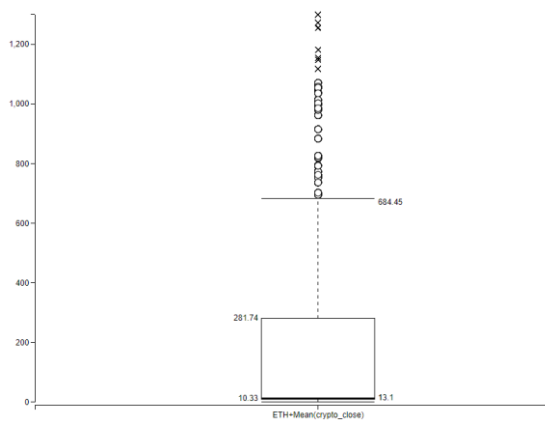
Boxplot da variação do preço do par *CHF/USD*



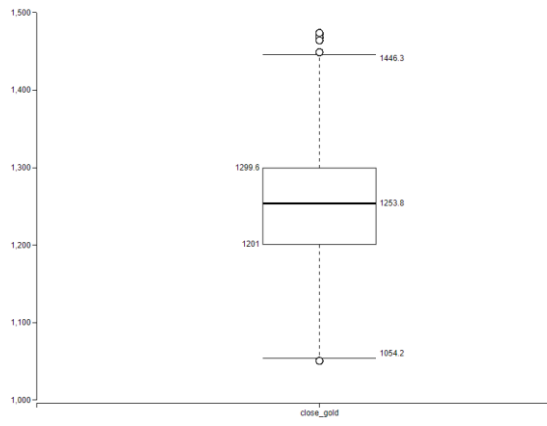
Boxplot da variação do preço da *BTC*



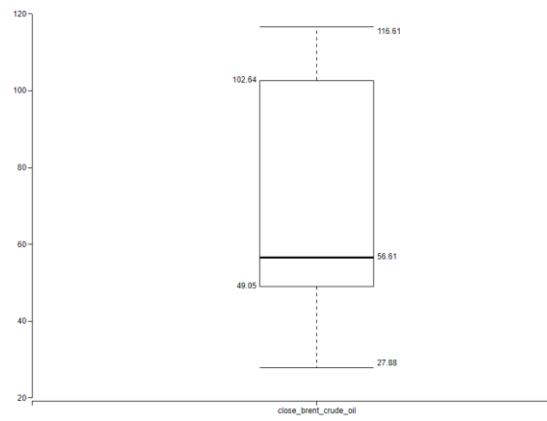
Boxplot da variação do preço da *LTC*



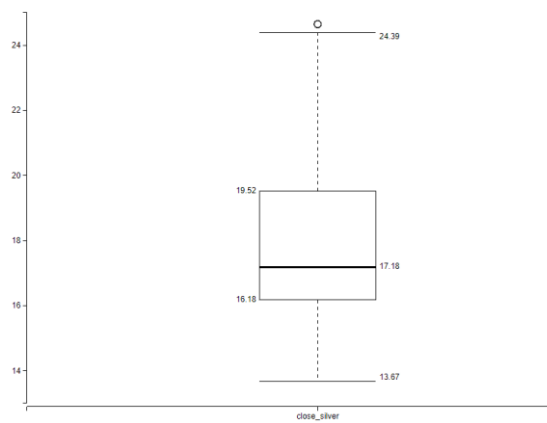
Boxplot da variação do preço da *ETH*



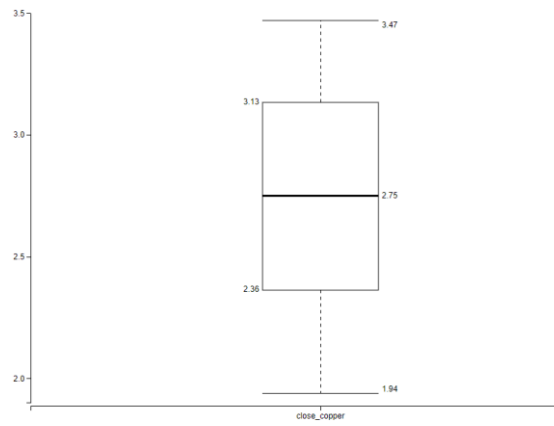
Boxplot da variação do preço do ouro



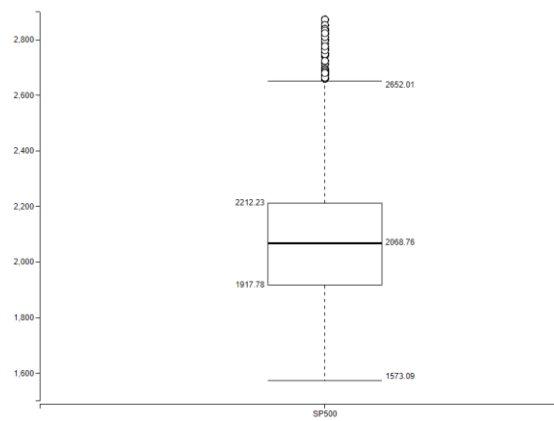
Boxplot da variação do preço do barril de petróleo



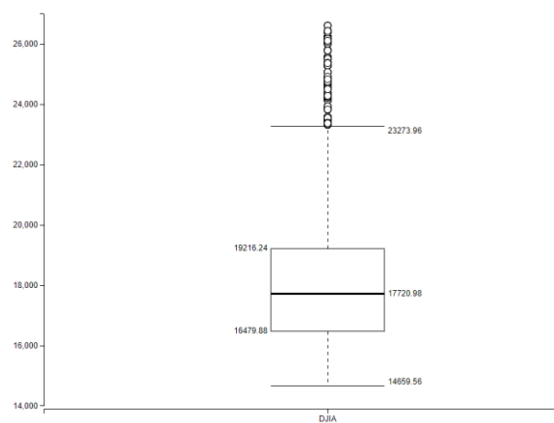
Boxplot da variação do preço da prata



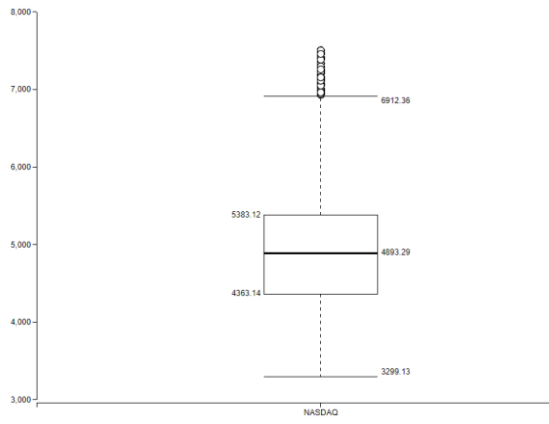
Boxplot da variação do preço do cobre



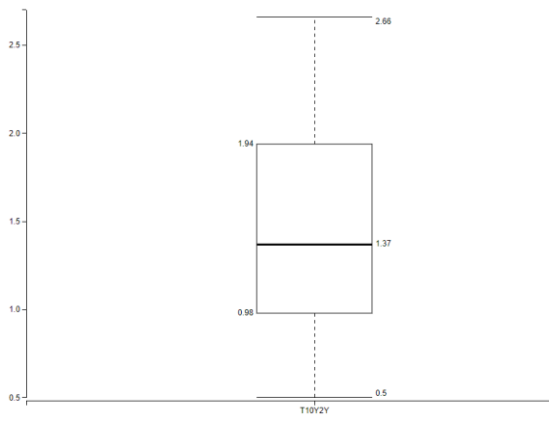
Boxplot da variação do preço do SP500



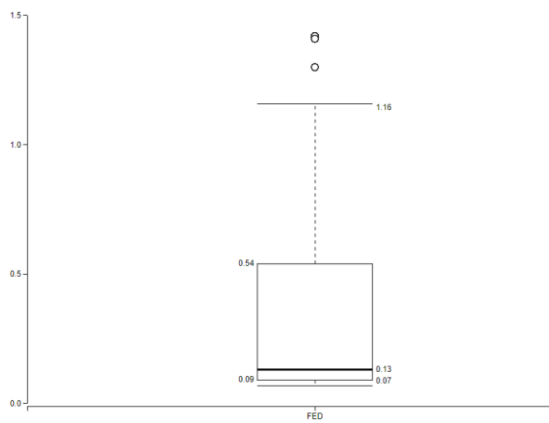
Boxplot da variação do preço do DJIA



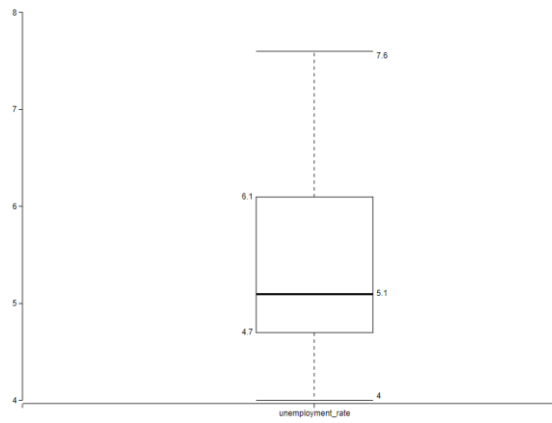
Boxplot da variação do preço do *NASDAQ*



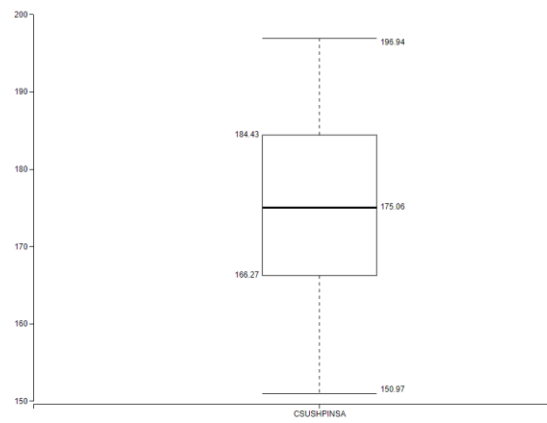
Boxplot da variação do preço do índice *T10Y2Y*



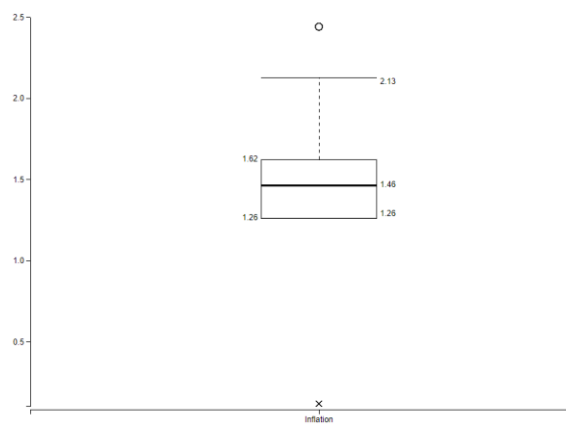
Boxplot da variação das taxas de juro do *FED*



Boxplot da variação da taxa de desemprego

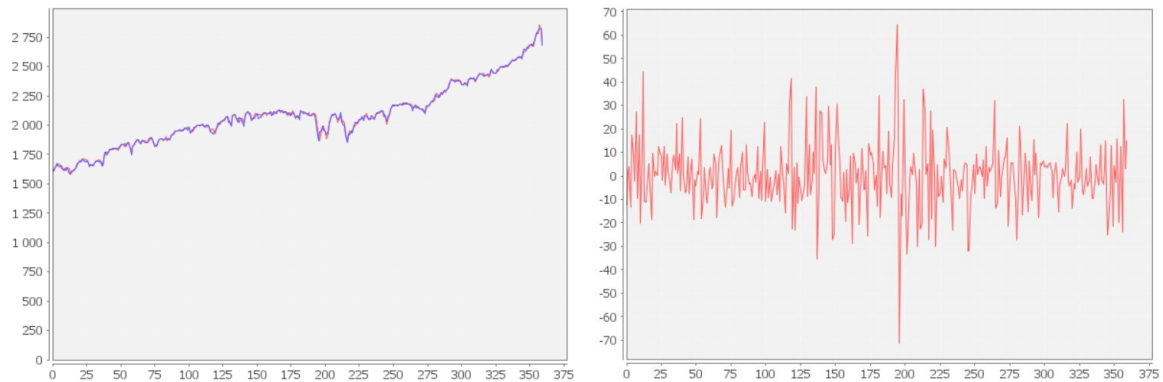


Boxplot da variação do índice *CSUSHPINSA*

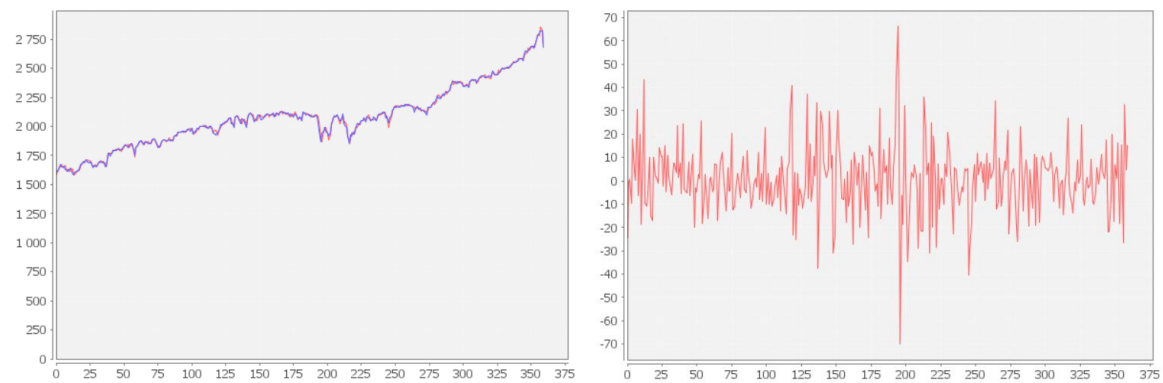


Boxplot da variação da Inflação

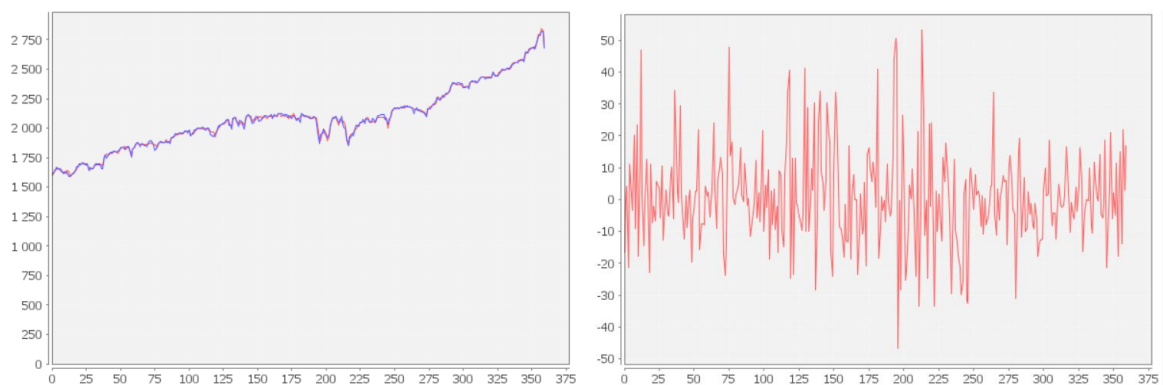
APÊNDICE B – GRÁFICOS DOS RESULTADOS SP500 (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



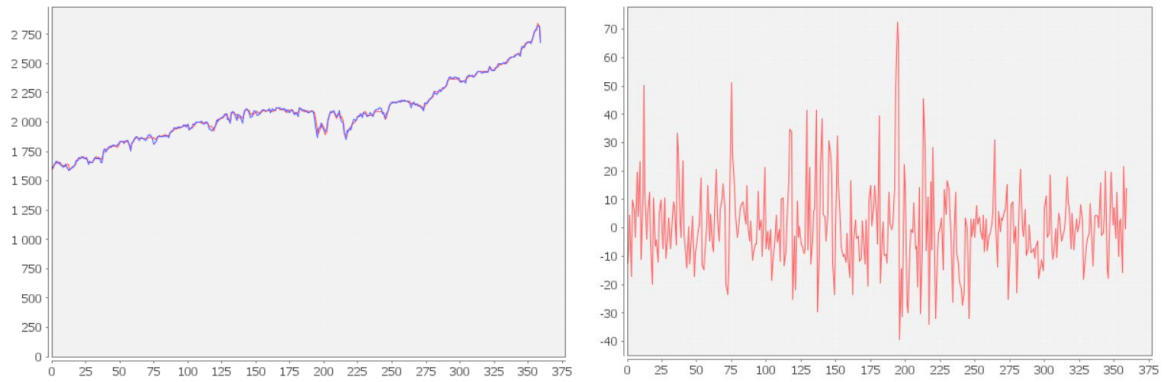
Linear Regression para o dia anterior à previsão



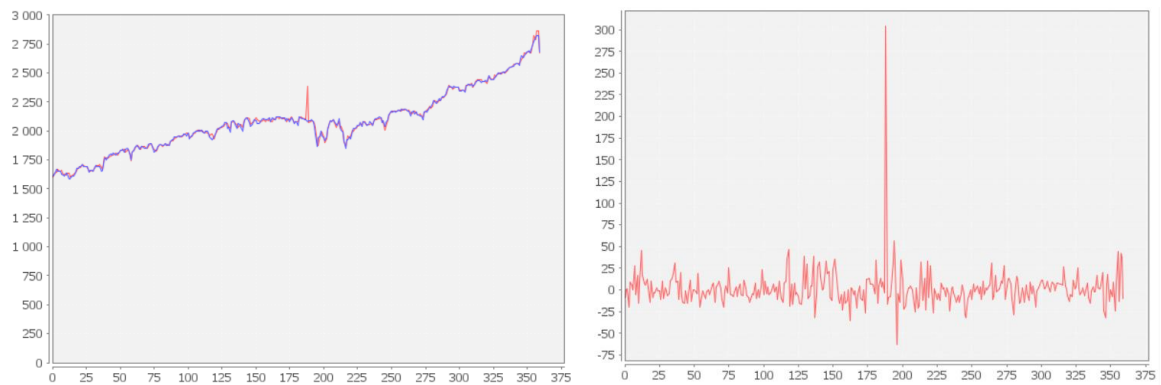
Linear Regression para os três dias anteriores à previsão



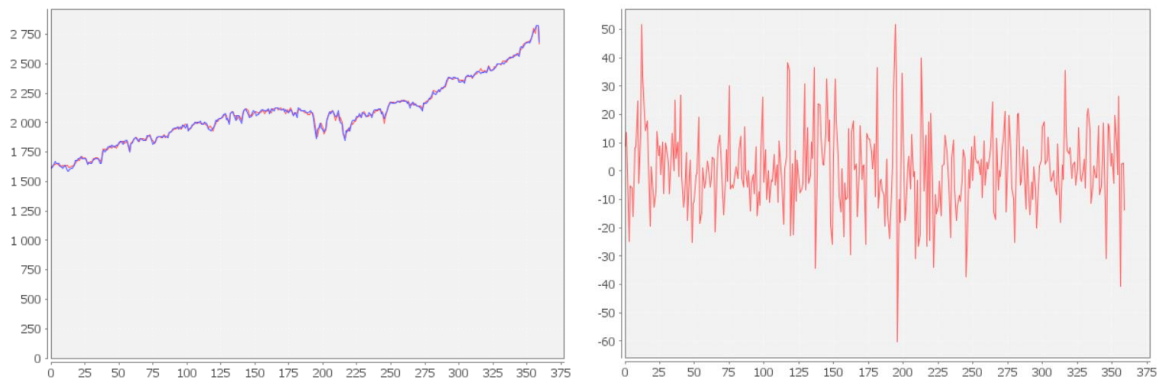
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

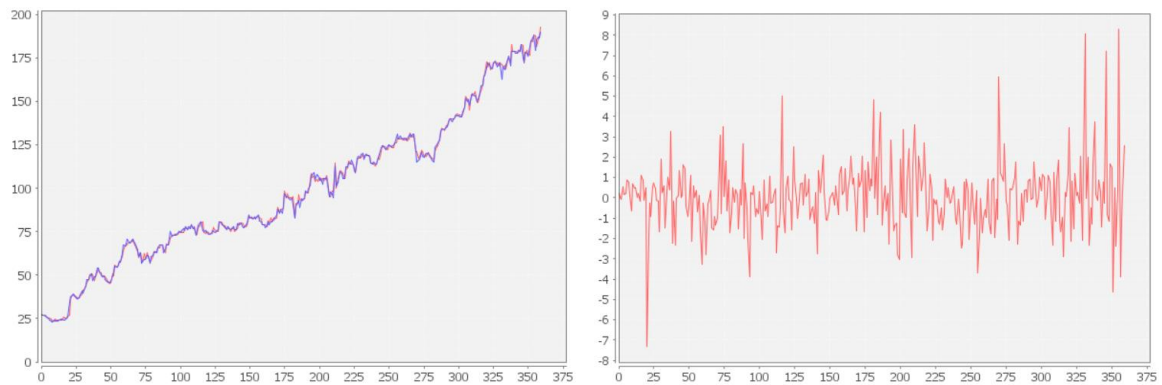


MultiLayer Perceptron para o dia anterior à previsão

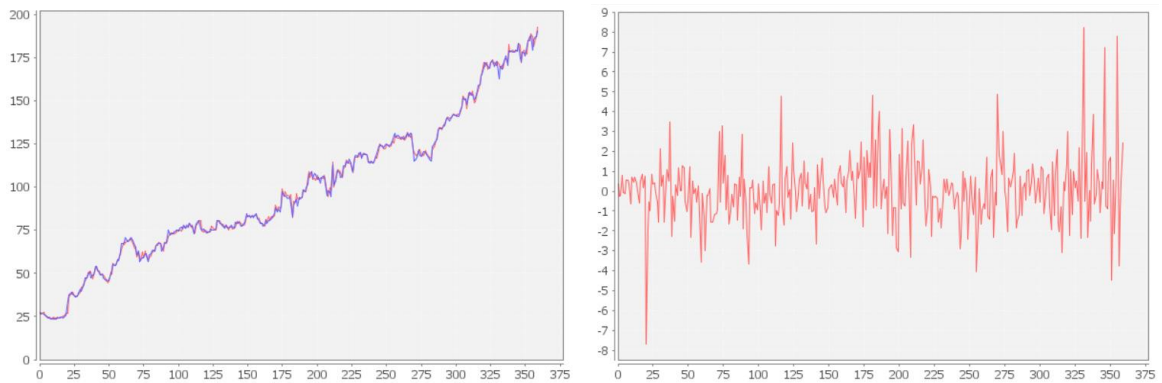


MultiLayer Perceptron para os três dias anteriores à previsão

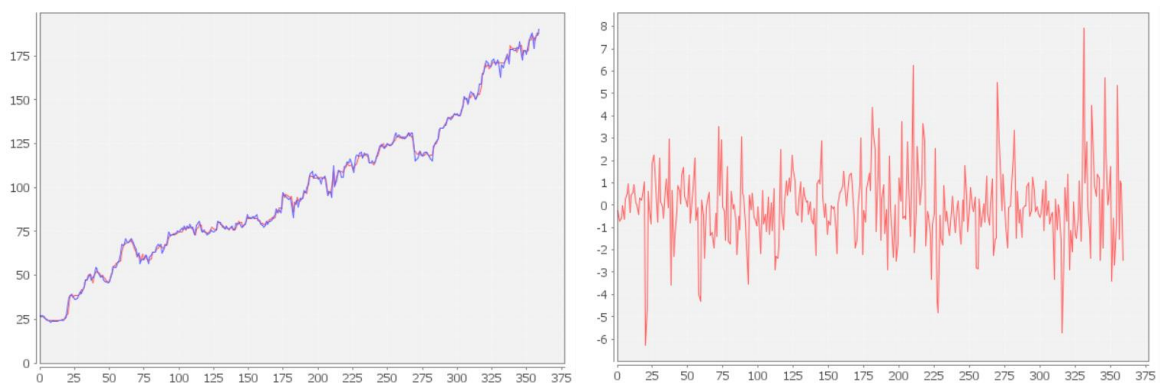
APÊNDICE C – GRÁFICOS DOS RESULTADOS FACEBOOK (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



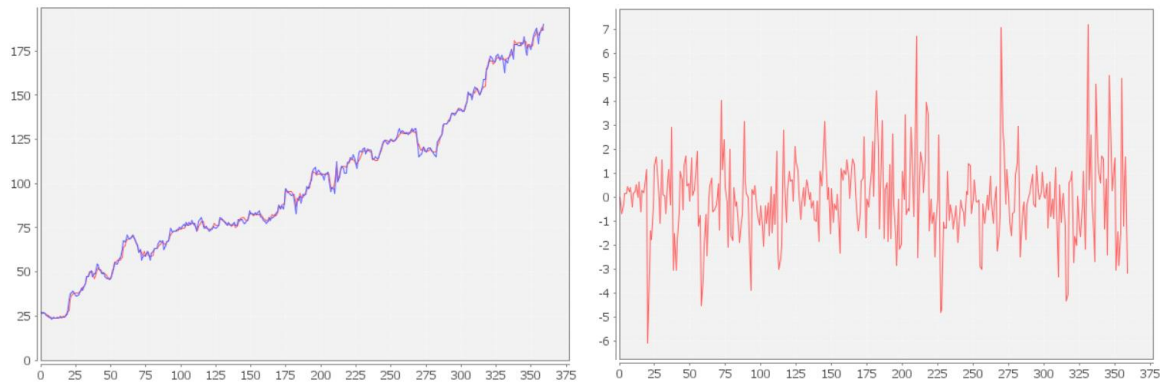
Linear Regression para o dia anterior à previsão



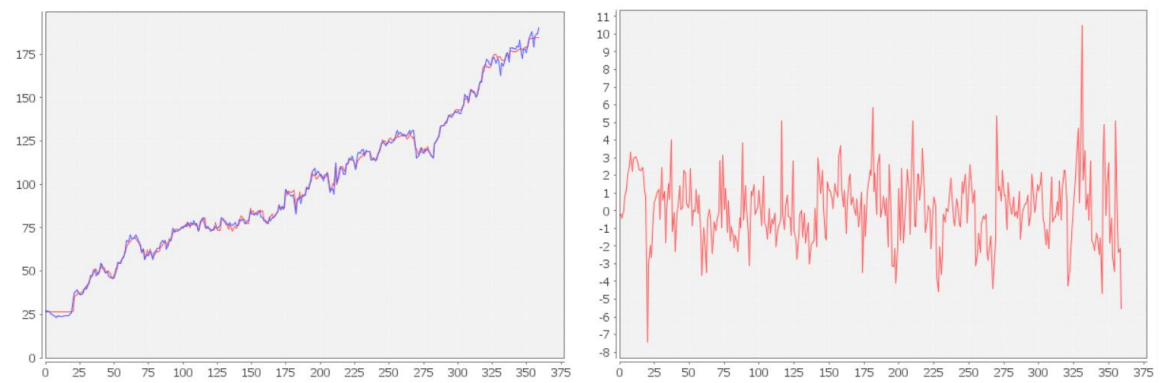
Linear Regression para os três dias anteriores à previsão



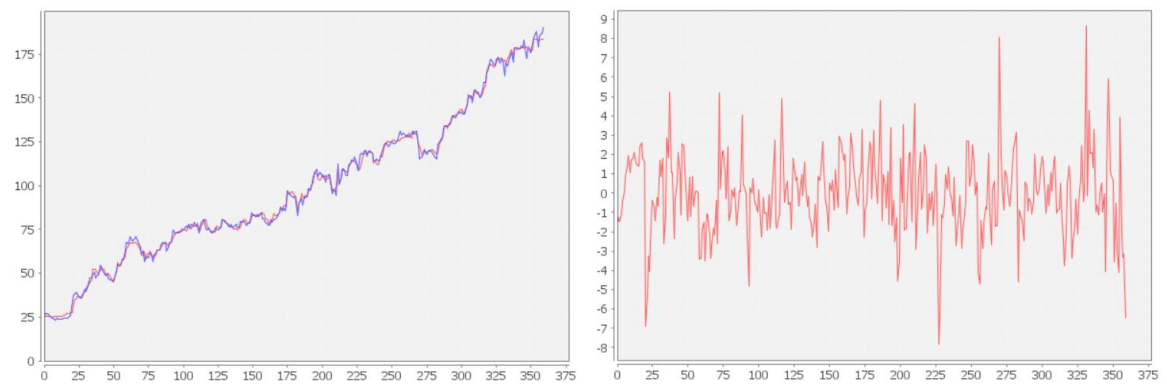
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

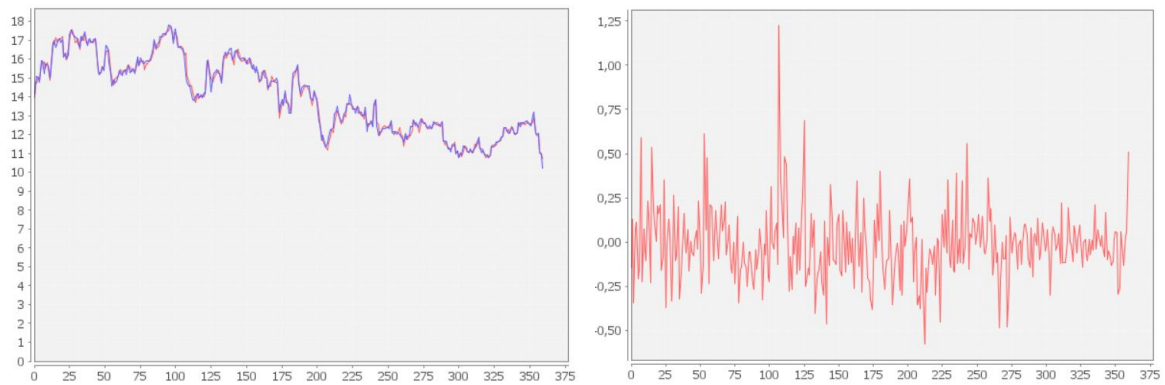


MultiLayer Perceptron para o dia anterior à previsão

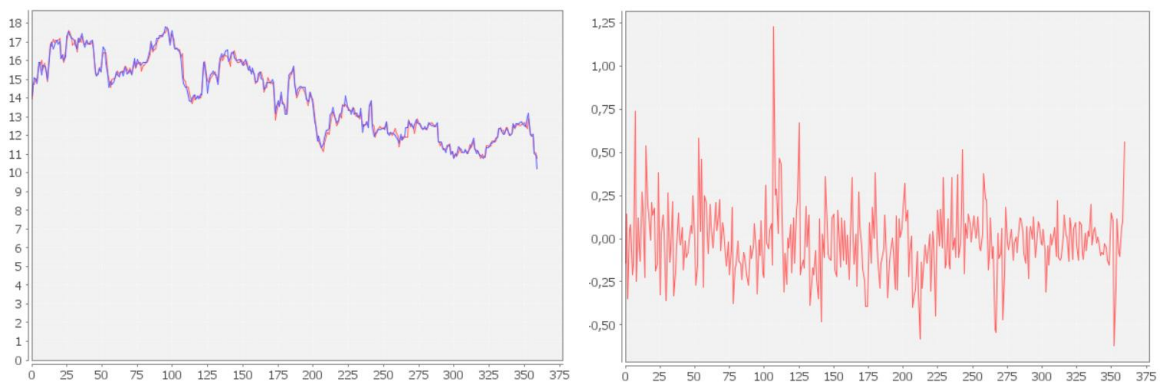


MultiLayer Perceptron para os três dias anteriores à previsão

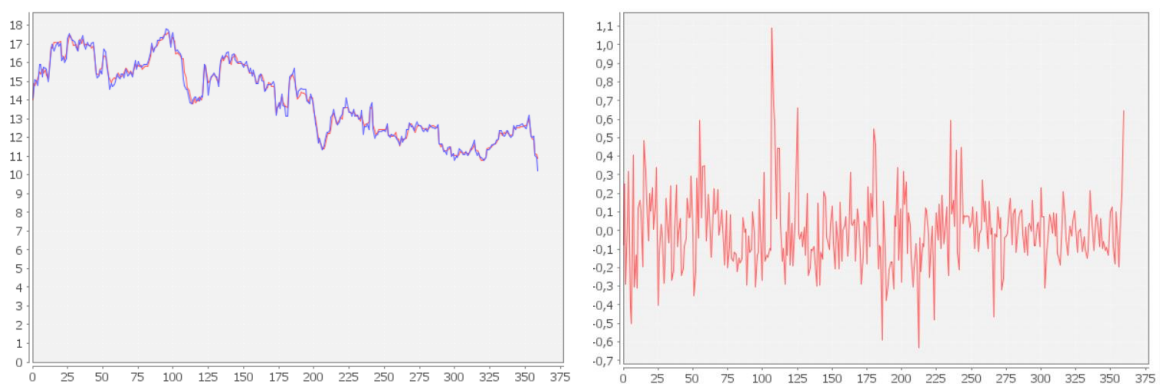
APÊNDICE D – GRÁFICOS DOS RESULTADOS FORD MOTOR COMPANY (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



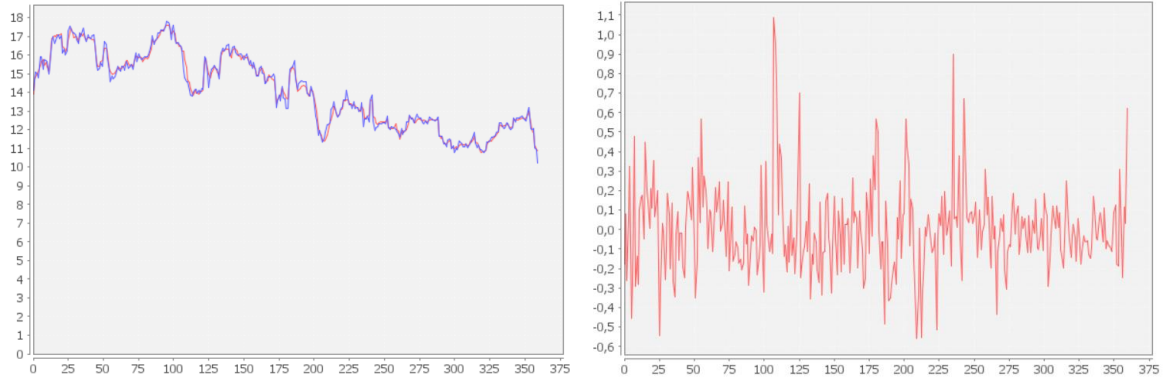
Linear Regression para o dia anterior à previsão



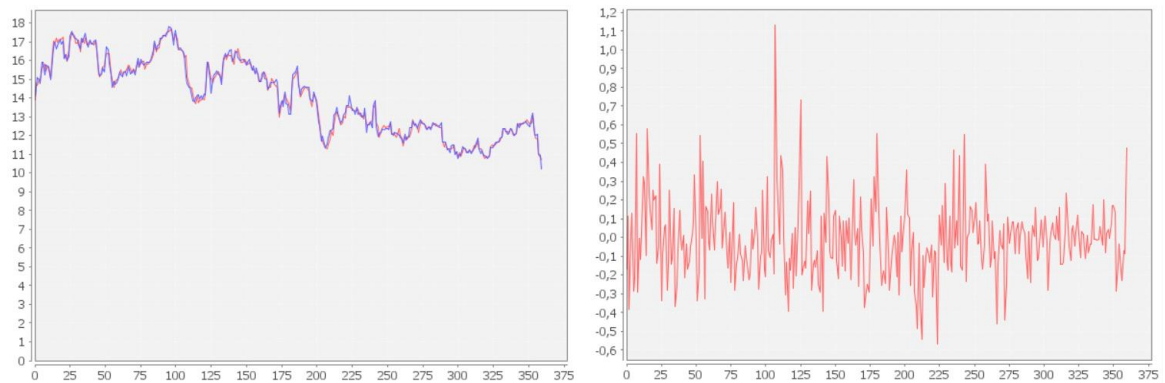
Linear Regression para os três dias anteriores à previsão



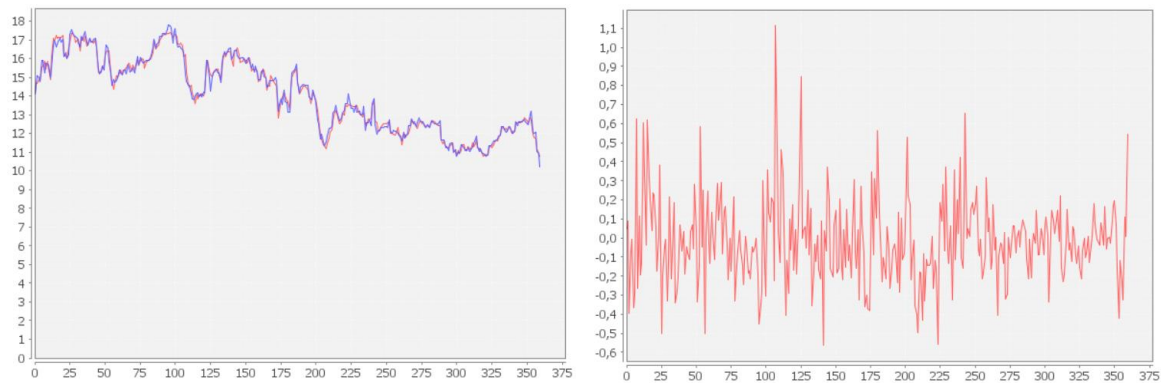
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

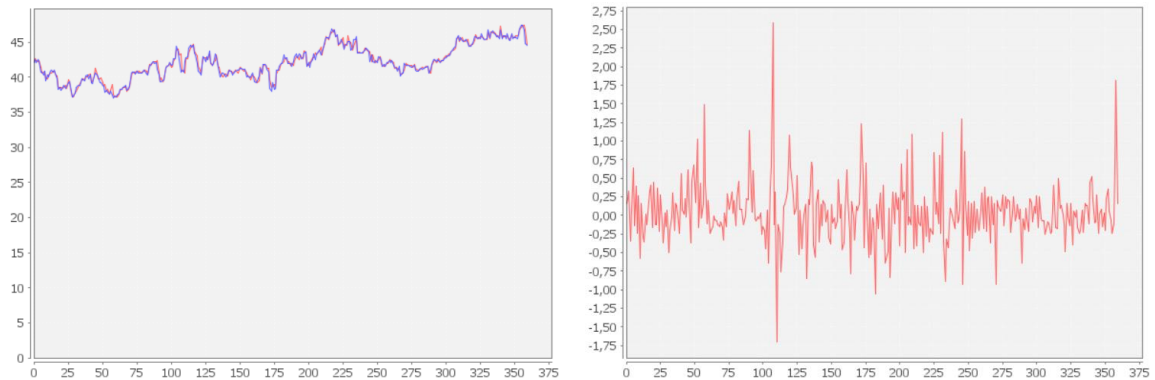


MultiLayer Perceptron para o dia anterior à previsão

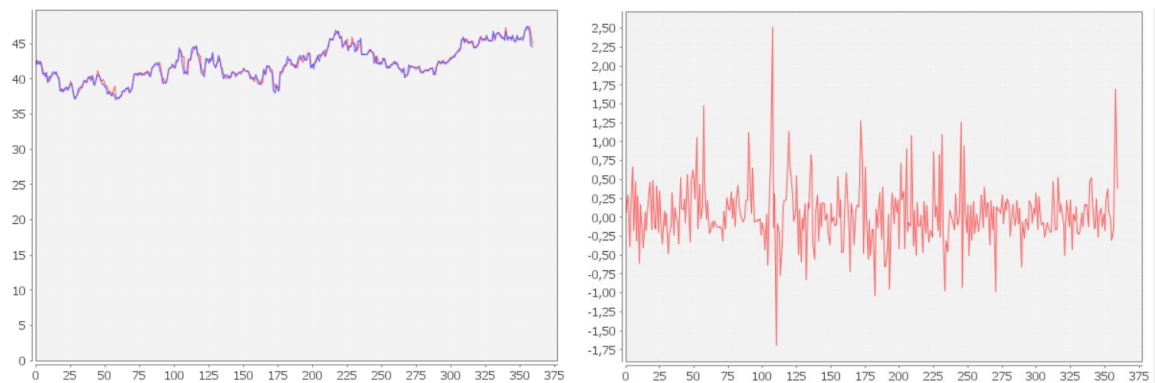


MultiLayer Perceptron para os três dias anteriores à previsão

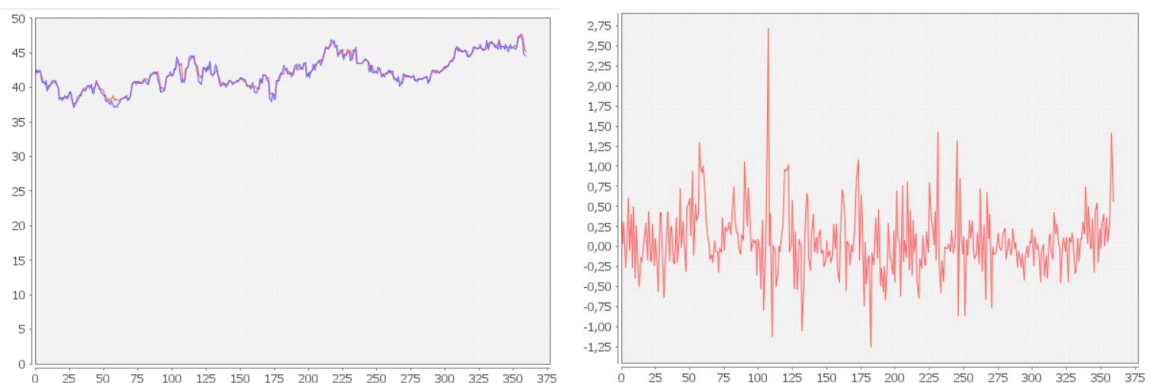
APÊNDICE E – GRÁFICOS DOS RESULTADOS COCA-COLA COMPANY (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



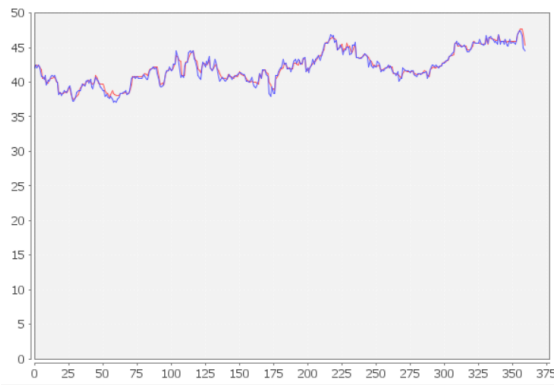
Linear Regression para o dia anterior à previsão



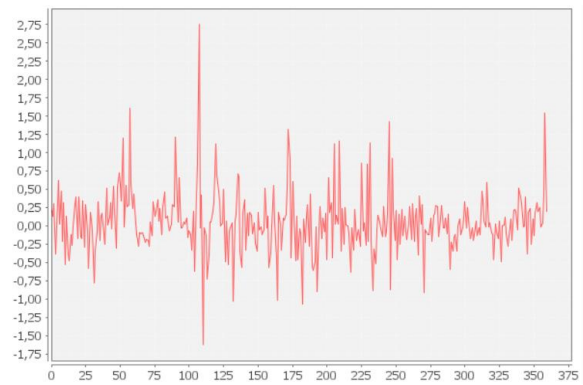
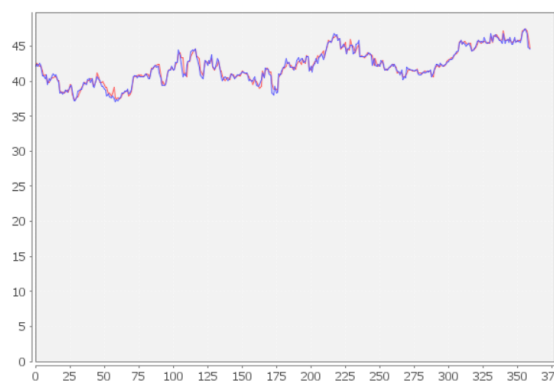
Linear Regression para os três dias anteriores à previsão



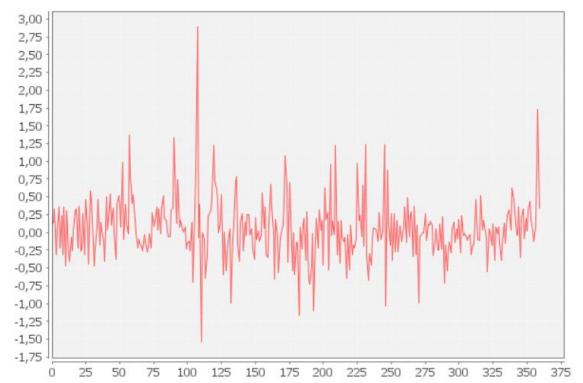
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

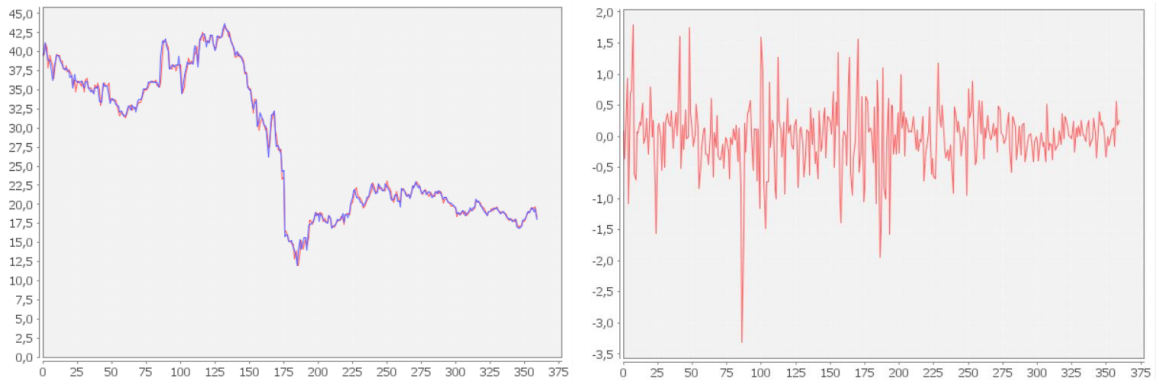


MultiLayer Perceptron para o dia anterior à previsão

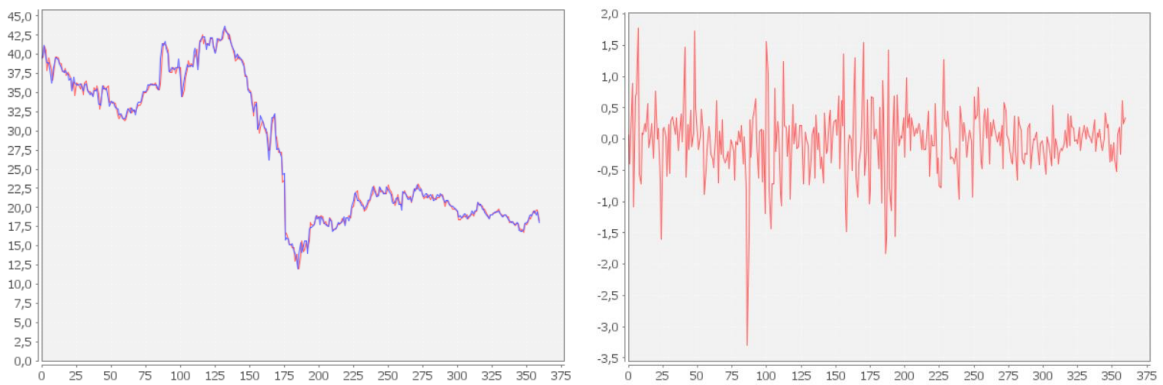


MultiLayer Perceptron para os três dias anteriores à previsão

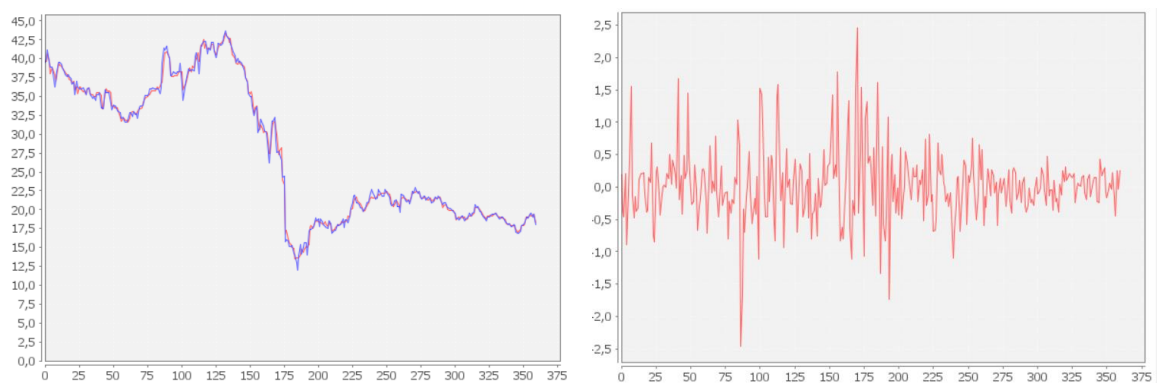
APÊNDICE F – GRÁFICOS DOS RESULTADOS KINDER MORGAN (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



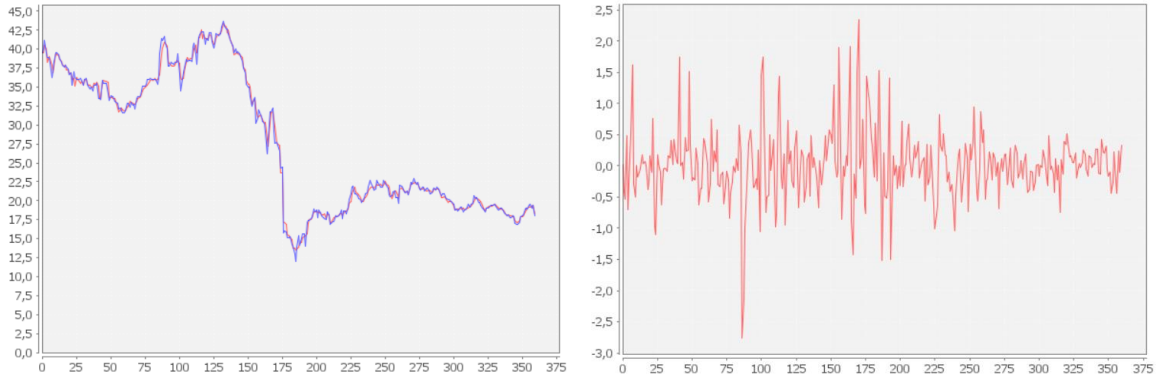
Linear Regression para o dia anterior à previsão



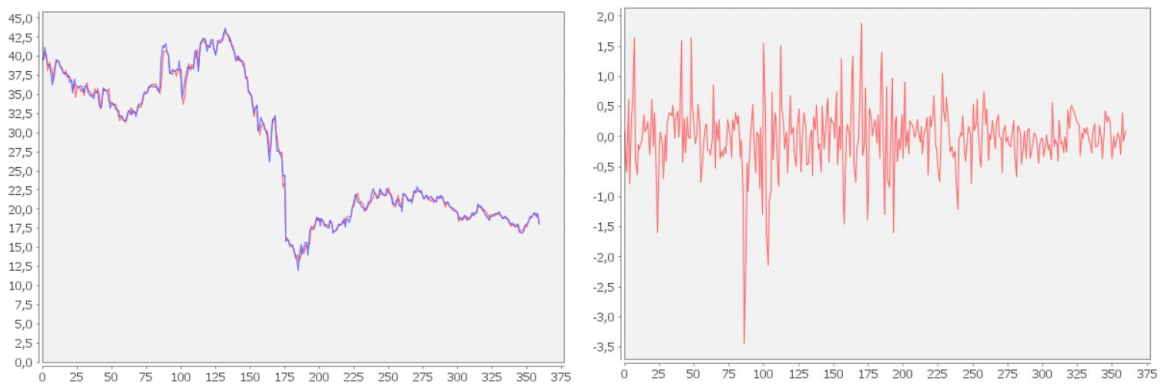
Linear Regression para os três dias anteriores à previsão



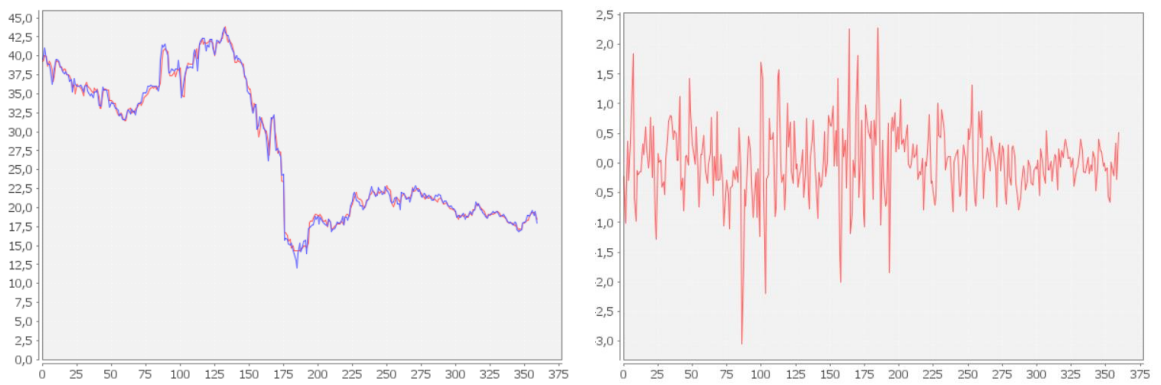
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

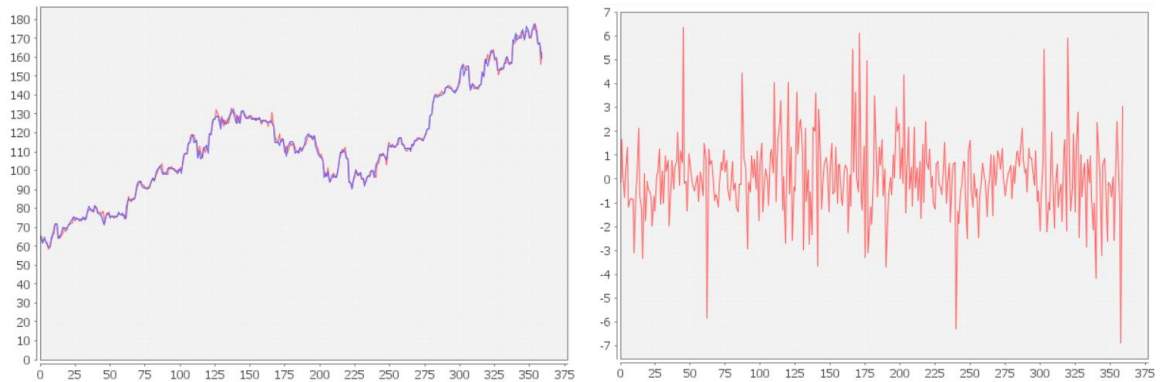


MultiLayer Perceptron para o dia anterior à previsão

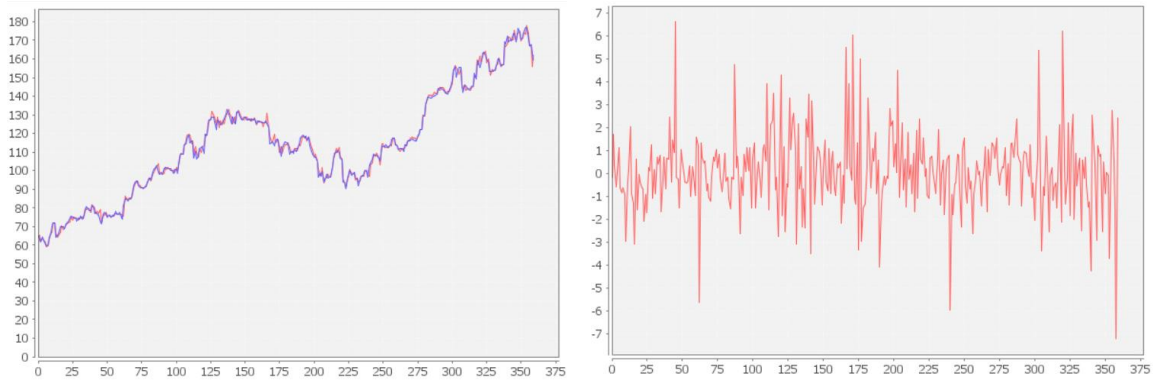


MultiLayer Perceptron para os três dias anteriores à previsão

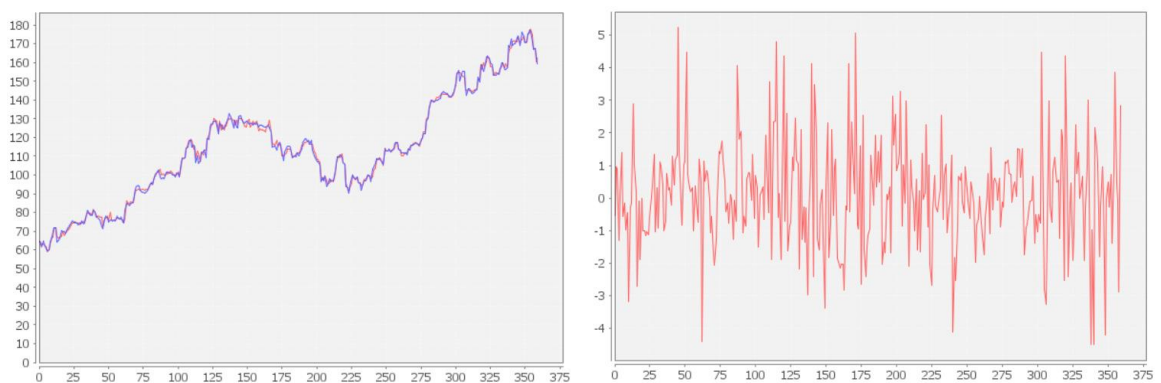
APÊNDICE G – GRÁFICOS DOS RESULTADOS APPLE (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



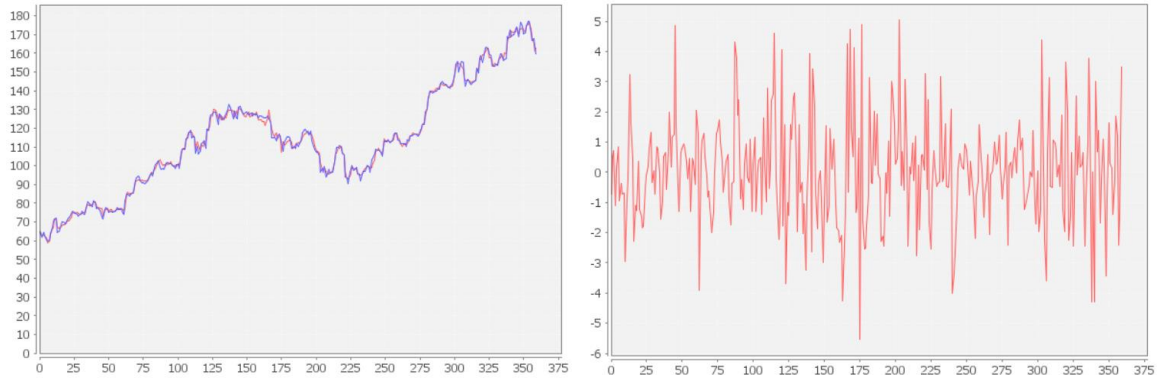
Linear Regression para o dia anterior à previsão



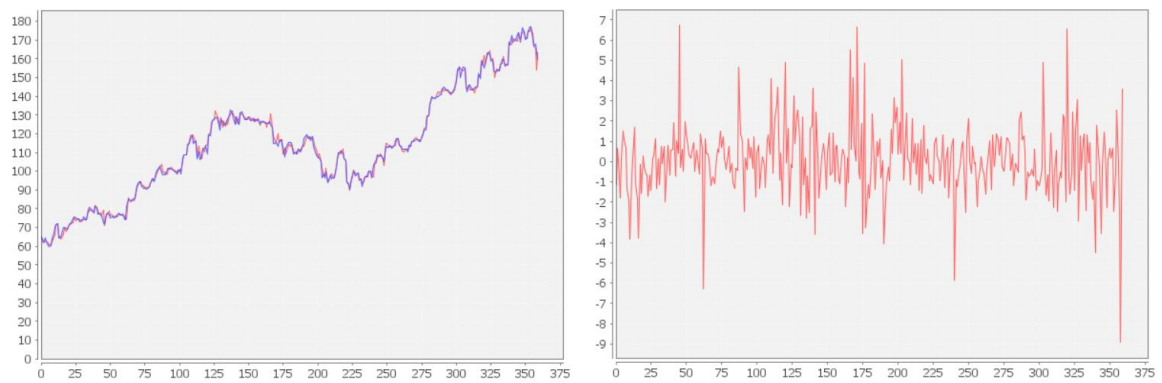
Linear Regression para os três dias anteriores à previsão



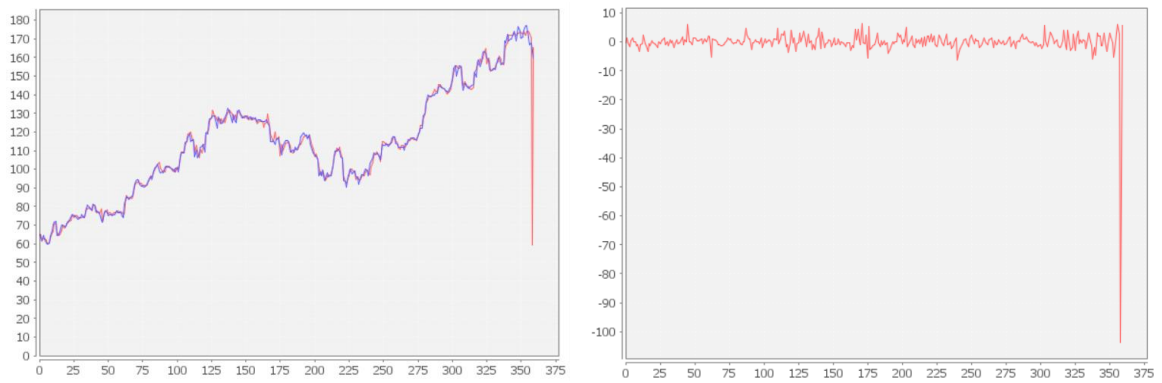
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

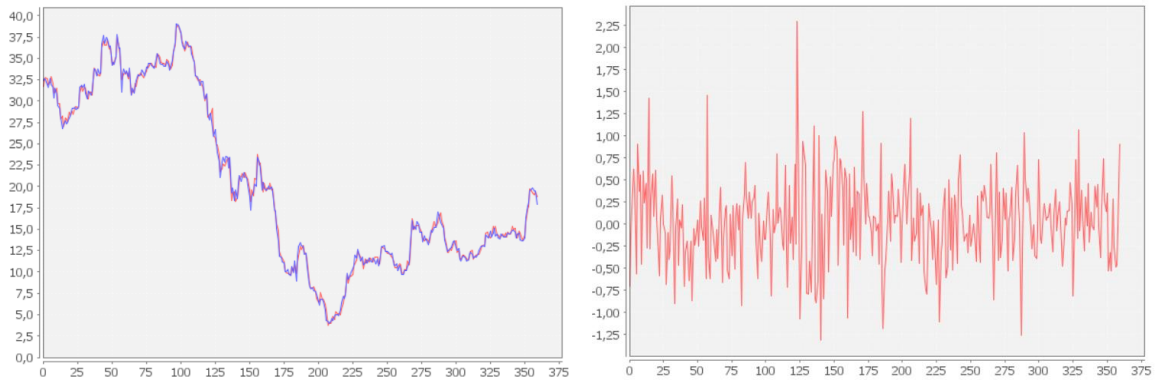


MultiLayer Perceptron para o dia anterior à previsão

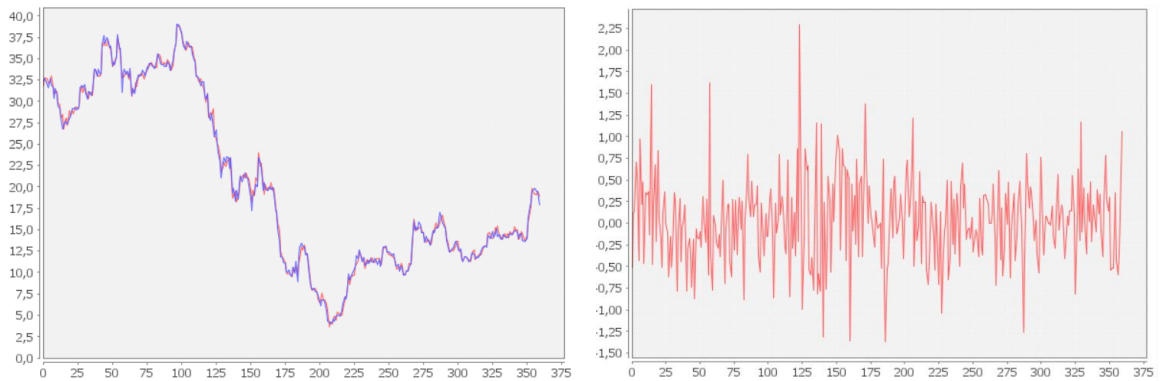


MultiLayer Perceptron para os três dias anteriores à previsão

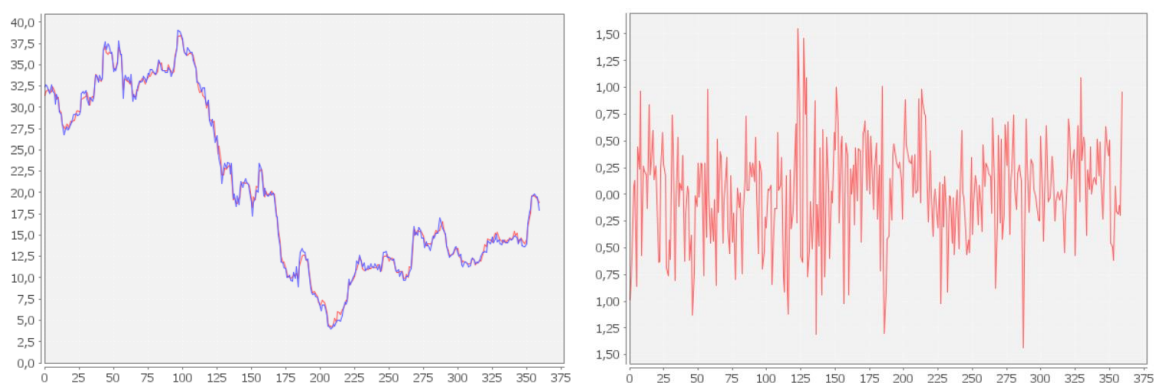
APÊNDICE H – GRÁFICOS DOS RESULTADOS FREEPORT-MCMORAN (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



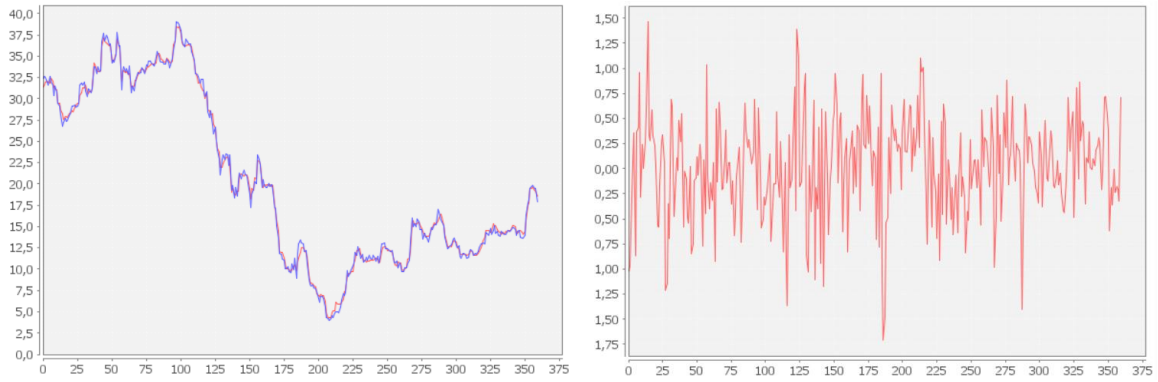
Linear Regression para o dia anterior à previsão



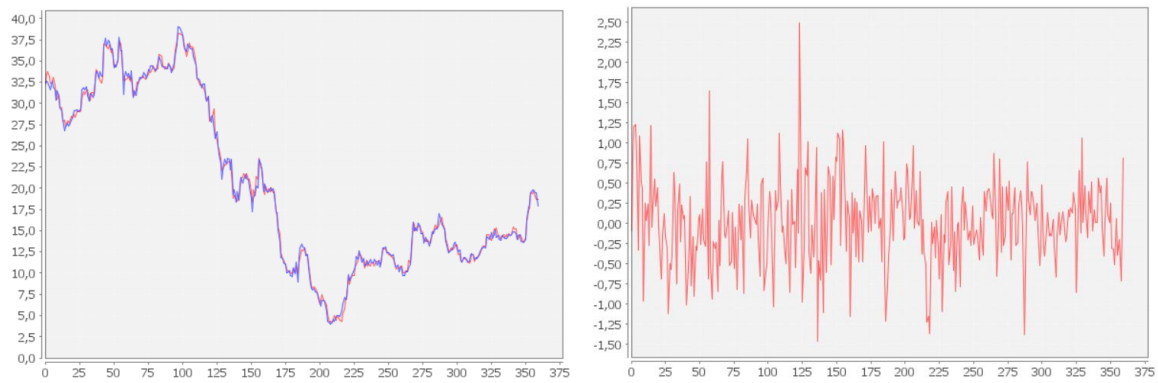
Linear Regression para os três dias anteriores à previsão



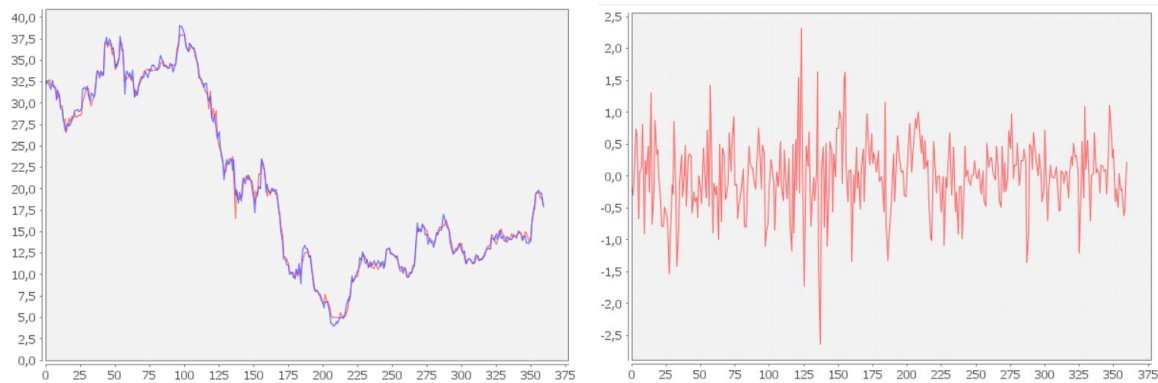
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

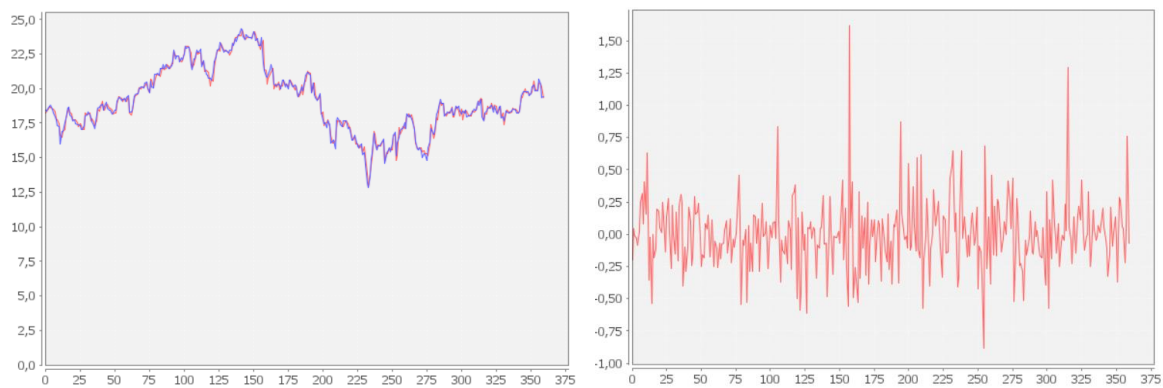


MultiLayer Perceptron para o dia anterior à previsão

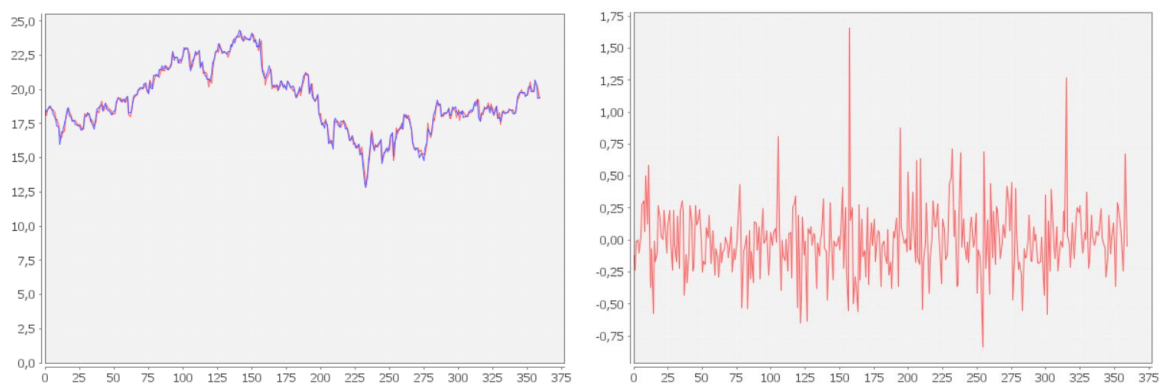


MultiLayer Perceptron para os três dias anteriores à previsão

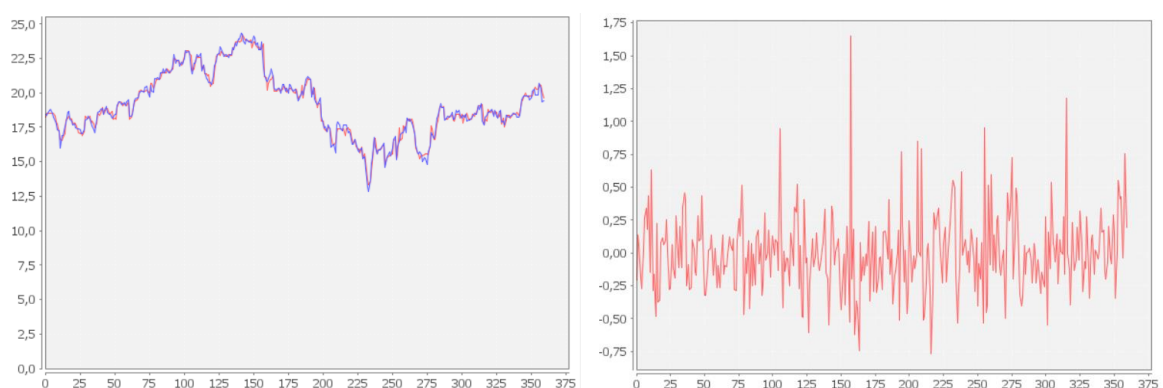
APÊNDICE I – GRÁFICOS DOS RESULTADOS HOST HOTELS & RESORTS (REAL VS PREVISÃO E DIFERENÇA ENTRE AMBOS)



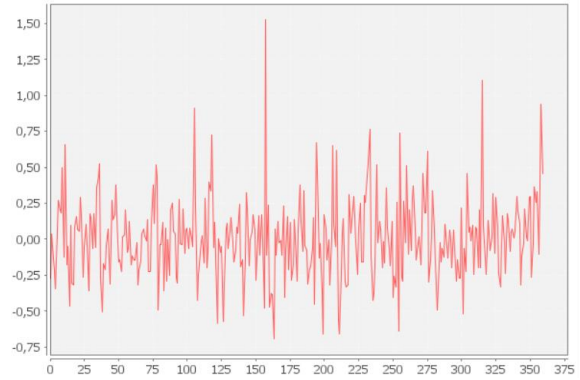
Linear Regression para o dia anterior à previsão



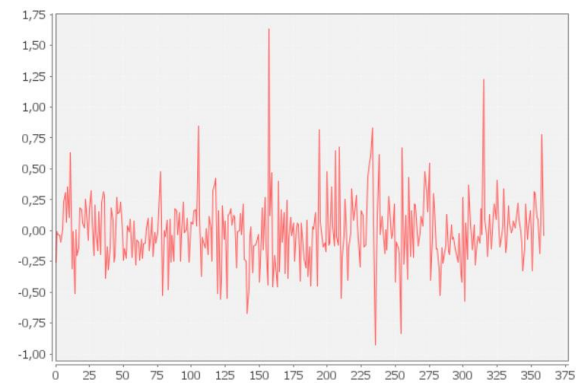
Linear Regression para os três dias anteriores à previsão



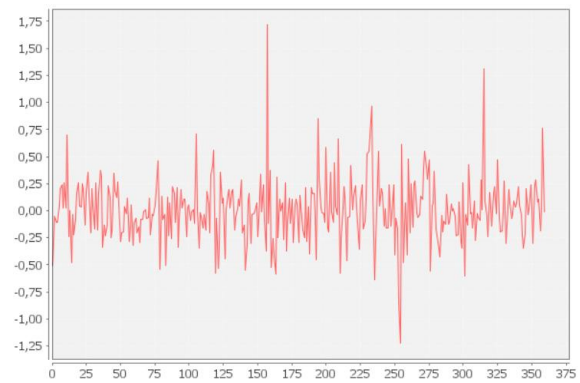
Random Forest Regression para o dia anterior à previsão



Random Forest Regression para os três dias anteriores à previsão

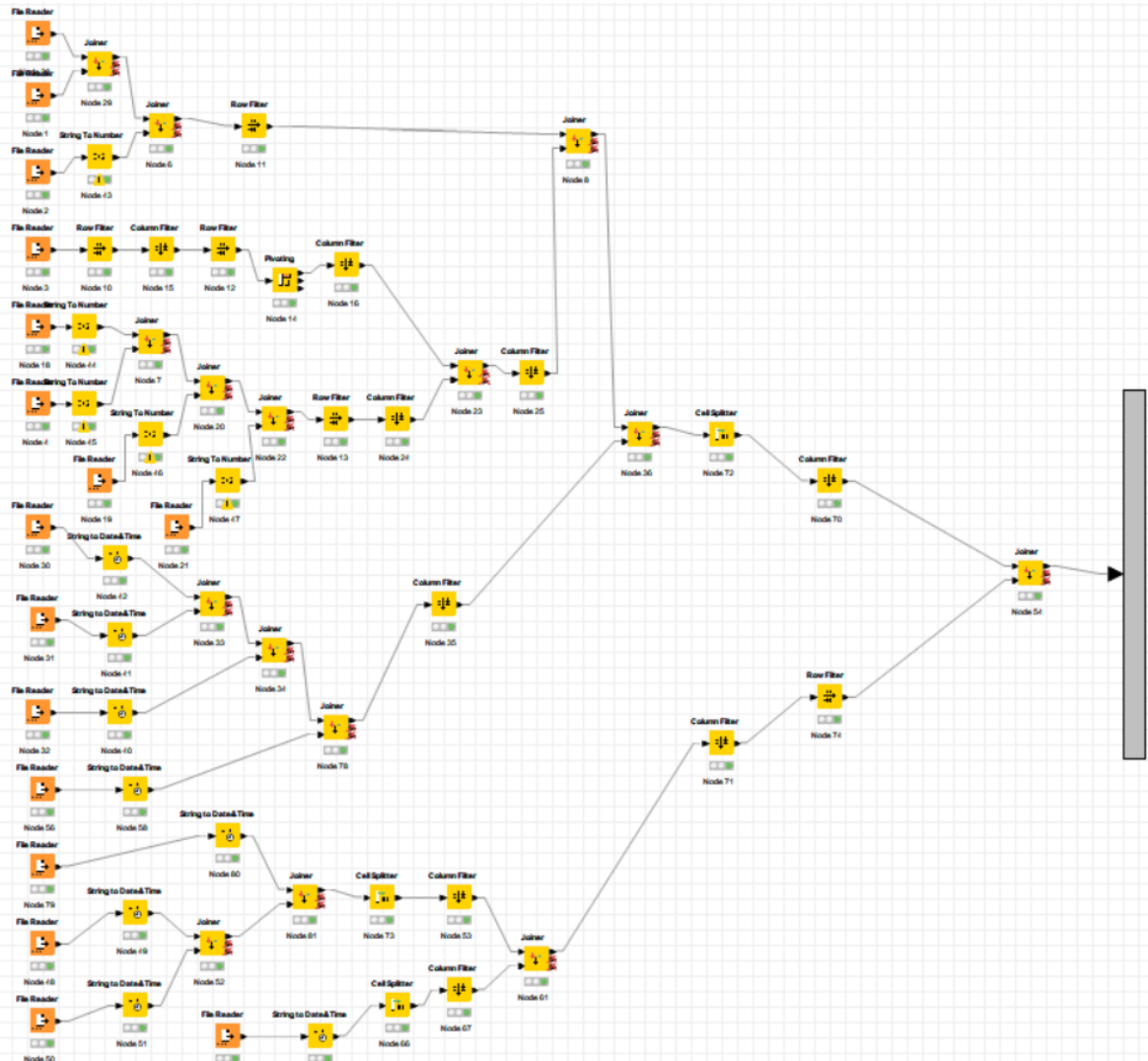


MultiLayer Perceptron para o dia anterior à previsão

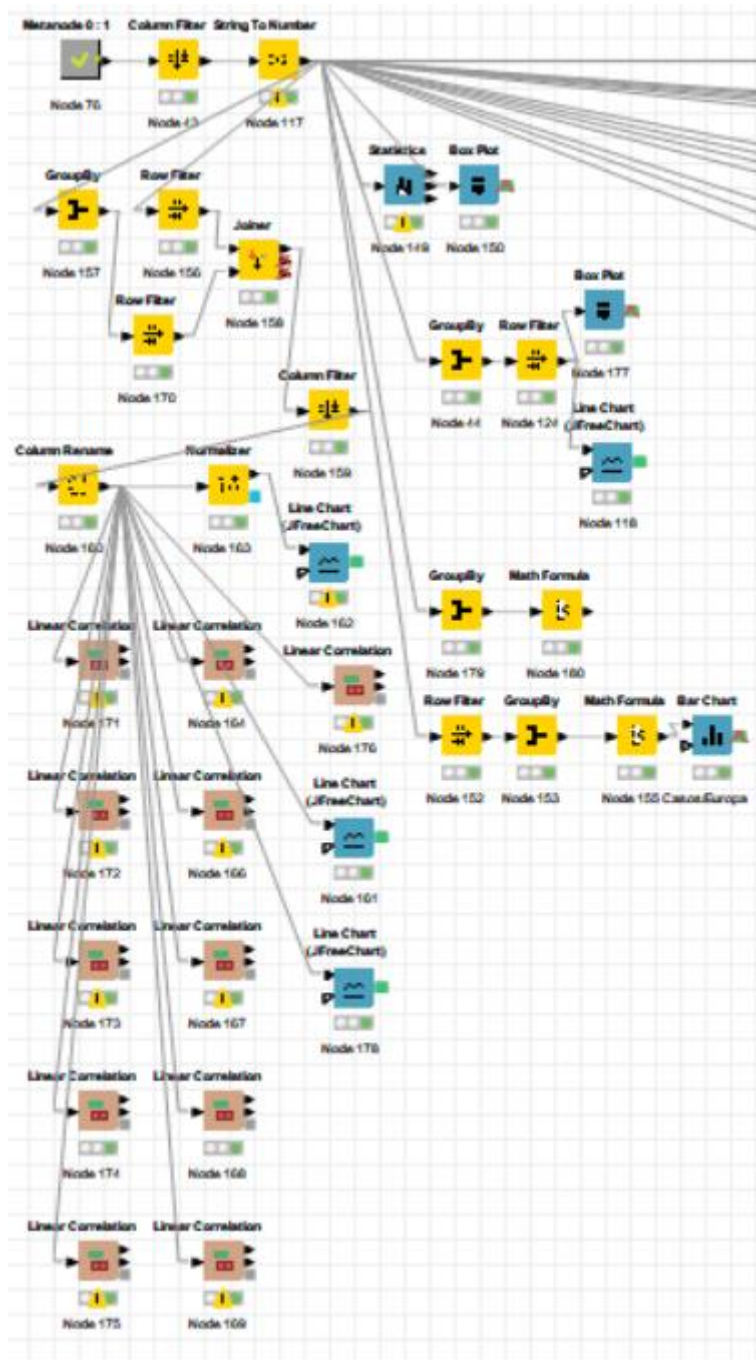


MultiLayer Perceptron para os três dias anteriores à previsão

APÊNDICE J – KNIME (INTEGRAÇÃO DAS BASES DE DADOS)



APÊNDICE K – KNIME (DATA UNDERSTANDING)



APÊNDICE L – KNIME (DATA PREPARATION)

