



Machine Learning Aplicado à Teoria de Portefólio

GONÇALO MENESES DE SOUSA

setembro de 2022

MACHINE LEARNING APLICADO À TEORIA DE PORTEFÓLIO

Gonçalo Meneses de Sousa

2022

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

MACHINE LEARNING APLICADO À TEORIA DE PORTEFÓLIO

Gonçalo Meneses de Sousa

Estudante n.º 1170476

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 e coorientação do Doutor Carlos Jorge Pereira Freitas.

2022

Instituto Superior de Engenharia do Porto

Departamento de Engenharia Mecânica

isen

P.PORTO

AGRADECIMENTOS

Em primeiro lugar, quero agradecer ao Diretor de Curso, Doutor Manuel Pereira Lopes, pela oportunidade que proporcionou aos alunos do mestrado de aprender de uma forma bastante prática que contribui para o nosso sucesso empresarial no curto prazo. Além disso, pelo desejo de integrar no plano de estudo unidades curriculares com uma visão para o futuro e que serviram para abrir os horizontes dos alunos.

Agradeço também ao Doutor Carlos Manuel Abreu Gomes Ferreira e ao Doutor Carlos Jorge Pereira Freitas por terem aceitado ser meu orientador e coorientador, respetivamente, e pela respetiva orientação, profissionalismo e apoio que sempre demonstraram ao longo do desenvolvimento da dissertação.

Aos meus pais, por todo o apoio e carinho que me deram ao longo de todo este percurso académico e que me permitiu alcançar o grau de Mestre em Engenharia e Gestão Industrial e assim concretizar mais um objetivo ambicionado. Ao meu falecido avô, por me ter incentivado a prosseguir os estudos no ramo das Engenharias e que sempre teve o sonho de ter um neto engenheiro tal como outrora foi.

Além do mais, quero agradecer aos meus colegas que me acompanharam nestes dois anos de mestrado e que sempre se mostraram disponíveis para me ajudar e motivar, desempenhando assim um papel fulcral durante o percurso académico. De realçar também todos os meus colegas que acompanharam todo este meu percurso académico durante os últimos 5 anos e que contribuíram ativamente no sucesso desta etapa estudantil.

Por último, agradeço à SEGIN, a empresa em que me encontro a trabalhar atualmente e que cooperou com a minha vida estudantil, sem nunca colocar qualquer entrave para o meu sucesso académico.

Agora novos desafios vão surgir e conto com o apoio de todos para alcançar todas as metas planeadas e almejadas.

página propositadamente em branco

RESUMO

Ao longo da vida é normal vivenciarmos altos e baixos a nível financeiro, mas necessitamos de estar preparados para ultrapassar estas situações da forma mais estável possível. Não podemos ficar dependentes de outras pessoas ou entidades para garantirmos o nosso futuro, é fundamental que as pessoas consigam desde cedo ter um pensamento a longo prazo em todos os aspetos, incluindo a nível financeiro. A literacia financeira é cada vez mais uma temática fundamental para a sociedade em geral, é fulcral que a pessoas sejam dotada de conhecimento e capacidade de gerir os seus ativos e posteriormente potenciar a sua saúde financeira.

Uma das formas de rentabilizar os ativos é efetuando investimentos, este podem ser feitos de diversas formas, contudo nesta dissertação vamos dar um especial enfoque nos investimentos no mercado financeiro, mais especificamente através de ETF's, *Commodities* e Fundos de Investimento. Assim, o principal objetivo da dissertação consistiu na aplicação de *Machine Learning* na Teoria de Portefólio para obtenção de carteira eficiente. Para a definição desta carteira de ativos eficiente foi necessário perceber a importância da diversificação do portefólio, de forma que os investidores ficassem menos suscetíveis a variações do mercado. Deste modo, foi dada especial importância ao risco e ao lucro na forma de atributos da carteira de ativos, pois foram estes que auxiliaram na determinação de um portefólio eficiente.

Numa primeira fase, foi efetuada uma análise dos diferente ativos para depois serem desenvolvidos modelos de previsão com base no histórico dos ativos e noutras variáveis que se correlacionavam com o conjunto de dados. Visto que o objetivo era maximizar os lucros e minimizar o risco, foram integrados na carteira diferentes tipos de investimentos, com diferentes riscos associados. Posteriormente, foi estudado o modelo que otimiza a determinação dos pesos dos ativos constituintes da carteira. Nesta última fase efetuou-se uma comparação entre o método AHP e o método de divisão igualitária das percentagens dos ativos na carteira de investimentos. Através da qual foi possível verificar que o método AHP se definia como o mais eficiente e flexível para a determinação da constituição da carteira de ativos, por possibilitar também aos investidores definir uma carteira mais conservadora ou mais especulativa.

De uma forma geral, este projeto pode ser útil para as indústrias de fundos de pensões, bancos, fundos de investimentos e entre outros. Para o desenvolvimento deste projeto foi utilizada a metodologia CRISP-DM e a Teoria do Portefólio e foram também utilizadas como ferramentas, o Knime e o Excel.

PALAVRAS-CHAVE

CRISP-DM, Teoria do Portefólio, Rentabilidade, Risco, Carteira de Ativos, Otimização

página propositadamente em branco

ABSTRACT

Throughout life, it is normal to experience ups and downs at a financial level, but we need to be prepared to overcome these situations in the most stable way possible. We cannot be dependent on other people or entities to ensure our future, people must be able to have long-term thinking from an early age in all aspects, including financially. Financial literacy is increasingly a key issue for society in general, people must be endowed with knowledge and ability to manage their assets and subsequently enhance their financial health.

One of the ways to monetize the assets is making investments, these can be made in several ways, however, in this dissertation we will give a special focus on financial market investments, more specifically through ETFs, Commodities and Investment Funds. Thus, the main goal of the dissertation was the application of Machine Learning in the Portfolio Theory to obtain an efficient portfolio. To define this efficient portfolio of assets it was necessary to understand the importance of portfolio diversification so that investors would be less susceptible to market variations. Therefore, special importance was given to risk and profit in the form of attributes of the portfolio of assets, since these were what helped in the determination of an efficient portfolio.

In the first phase, an analysis of the different assets was carried out and then forecasting models were developed based on asset history and other variables that correlated with the data set. Since the objective was to maximize profits and minimize risk, different types of investments with different associated risks were integrated into the portfolio. Subsequently, the model that optimises the determination of the weights of the assets that make up the portfolio was studied. In this last phase, a comparison was made between the AHP method and the method of equal division of asset percentages in the investment portfolio. Through this comparison, it was possible to verify that the AHP method was defined as the most efficient and flexible method to determine the constitution of the portfolio of assets since it also allows investors to define a more conservative or more speculative portfolio.

In general, this project can be useful for the industries of pension funds, banks, investment funds and among others. To develop this project we used the CRISP-DM methodology and the Portfolio Theory, and we also used Knime and Excel as tools.

KEYWORDS

CRISP-DM, Portfolio theory, Profitability, Risk, Asset portfolio, Optimization

página propositadamente em branco

ÍNDICE

| | |
|---|------|
| ÍNDICE DE FIGURAS | IX |
| ÍNDICE DE TABELAS | XI |
| LISTAS DE SIGLAS E SÍMBOLOS..... | XIII |
| 1. INTRODUÇÃO | 1 |
| 1.1. Problema de investigação, enquadramento e pertinência..... | 1 |
| 1.2. Questão e objetivos de investigação..... | 2 |
| 1.3. Opções metodológicas | 3 |
| 2. REVISÃO BIBLIOGRÁFICA..... | 5 |
| 2.1. CRISP-DM..... | 5 |
| 2.1.1. Business Understanding..... | 6 |
| 2.1.2. Data Understanding | 6 |
| 2.1.3. Data Preparation..... | 6 |
| 2.1.4. Modeling | 7 |
| 2.1.5. Evaluation..... | 8 |
| 2.1.6. Deployment..... | 9 |
| 2.2. Knime..... | 9 |
| 2.3. Rentabilidade e Risco das Carteiras | 10 |
| 2.4. Teoria do Portfólio | 15 |
| 2.5. Tipos de Investimentos Financeiros | 19 |
| 2.5.1. Ações..... | 19 |
| 2.5.2. Commodities | 20 |
| 2.5.3. Real Estate..... | 20 |
| 2.5.4. Criptomoedas..... | 20 |
| 2.6. Método AHP | 21 |
| 3. CASO DE ESTUDO | 23 |
| 3.1. Business Understanding | 23 |
| 3.2. Data Understanding | 24 |
| 3.2.1. Recolha de Dados..... | 24 |
| 3.2.2. Descrição dos Dados | 24 |
| 3.2.3. Exploração dos Dados | 26 |
| 3.3. Data Preparation | 35 |
| 3.3.1. Seleção de Dados | 35 |
| 3.3.2. Limpeza e Preparação de Dados | 35 |
| 3.4. Modeling | 36 |
| 3.4.1. Seleção das Técnicas de Modelação | 36 |
| 3.4.2. Modelo de Teste | 37 |
| 3.4.3. Construção de Modelos | 37 |

| | |
|---|----|
| 3.4.4. Avaliação de Modelos | 38 |
| 3.5. Otimização Carteira de Ativos | 40 |
| 4. CONCLUSÃO | 45 |
| 4.1. Conclusões finais | 45 |
| 4.2. Limitações e investigação futura | 45 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 47 |
| APÊNDICE A | 49 |
| APÊNDICE B | 50 |
| APÊNDICE C | 51 |
| APÊNDICE D | 52 |
| APÊNDICE E | 53 |
| APÊNDICE F | 54 |
| APÊNDICE G | 55 |
| APÊNDICE H | 56 |
| APÊNDICE I | 57 |

página propositadamente em branco

ÍNDICE DE FIGURAS

| | |
|--|----|
| Figura 1 – Modelo CRISP-DM (Colpani, 2019)..... | 5 |
| Figura 2 – Interface do software Knime (Knime, 2022) | 10 |
| Figura 3 – Evolução do risco da carteira relativamente ao nº de ações (Brealey, 2020)..... | 12 |
| Figura 4 – Riscos da carteira (Henri, 2018) | 13 |
| Figura 5 – Variância de uma carteira com duas ações (Brealey, 2020) | 13 |
| Figura 6 – Efeito da variação do beta nas carteiras (Brealey, 2020)..... | 14 |
| Figura 7 – Distribuição Normal de Investimento (Brealey, 2020)..... | 15 |
| Figura 8 – Gráfico de Rentabilidade e Risco de Carteiras (Brealey, 2020)..... | 16 |
| Figura 9 – Efeito dos empréstimos na carteira (Brealey, 2020)..... | 17 |
| Figura 10 – Gráfico descritivo do modelo CAPM (Brealey, 2020)..... | 18 |
| Figura 11 - Estrutura Hierárquica de Método AHP | 21 |
| Figura 12 – Variação relativa dos ativos do grupo S&P500..... | 27 |
| Figura 13 – Variação relativa acumulada dos ativos do grupo S&P500..... | 27 |
| Figura 14 – Matriz de correlação dos ativos do grupo S&P500 | 28 |
| Figura 15 – Variação relativa dos ativos do grupo ETF's..... | 29 |
| Figura 16 – Variação relativa acumulada dos ativos do grupo ETF's | 29 |
| Figura 17 – Matriz de correlação dos ativos do grupo ETF's..... | 31 |
| Figura 18 – Gráfico de menores correlações do grupo ETF's..... | 31 |
| Figura 19 – Variação relativa dos ativos do grupo Commodities..... | 32 |
| Figura 20 – Variação relativa acumulada dos ativos do grupo Commodities..... | 33 |
| Figura 21 – Matriz de correlação dos ativos do grupo Commodities | 34 |
| Figura 22 - Gráfico de menores correlações do grupo Commodities | 34 |
| Figura 23 – Resultados Otimização Carteira de Ativos através do Método AHP..... | 42 |
| Figura 24 – Resultados Otimização Carteira de Ativos através do Método de divisão igualitária .. | 43 |

página propositadamente em branco

ÍNDICE DE TABELAS

| | |
|---|----|
| Tabela 1 – Valor resultante do investimento | 10 |
| Tabela 2 – Taxa média de retorno anual..... | 11 |
| Tabela 3 – Variabilidade das carteiras..... | 12 |
| Tabela 4 – Carteiras eficientes escolhidas de dez ações..... | 16 |
| Tabela 5 – Atributos do Grupo ETF's | 24 |
| Tabela 6 – Atributos do Grupo S&P500 | 25 |
| Tabela 7 – Atributos do Grupo Commodities..... | 26 |
| Tabela 8 – Correlação entre os ativos do grupo S&P500..... | 28 |
| Tabela 9 – Maiores correlações entre os ativos do grupo ETF's..... | 30 |
| Tabela 10 – Maiores correlações entre os ativos do grupo Commodities..... | 33 |
| Tabela 11 - Avaliação Modelos Grupo ETF's | 38 |
| Tabela 12 – Avaliação Modelos Grupo Commodities | 39 |
| Tabela 13 – Avaliação Modelos ETFs..... | 40 |
| Tabela 14 – Previsão de Lucro e Risco dos Ativos..... | 41 |
| Tabela 15 - Previsão e Realidade de Lucro do mês em estudo..... | 43 |

página propositadamente em branco

LISTAS DE SIGLAS E SÍMBOLOS

Lista de Siglas

| | |
|----------|---|
| ISEP | Instituto Superior de Engenharia do Porto |
| P.Porto | Instituto Politécnico do Porto |
| CAPM | Capital Asset Pricing Model |
| S&P500 | Standard and Poor's 500 |
| CRISP-DM | Cross-Industry Standard Process for Data Mining |
| DCA | Dollar-cost averaging |
| APT | Teoria da avaliação por arbitragem |
| REIT | Real Estate Investment Trust |
| Knime | Konstanz Information Miner |
| AHP | Analytic Hierarchy Process |
| EQM | Erro Quadrático Médio |

Lista de Símbolos

| | | |
|------------|------------------------|------|
| r_m | Retorno do mercado | % |
| r_f | Taxa de juro sem risco | % |
| N | Número de observações | un |
| σ | Desvio padrão | % |
| σ^2 | Variância | % |
| β | Beta | un |

página propositadamente em branco

1. INTRODUÇÃO

Este capítulo divide-se em três tópicos sendo que primeiramente é abordado o problema de investigação, enquadramento e pertinência, no qual é descrita a situação e são analisadas as principais causas do problema. Posteriormente, na questão e objetivos de investigação são discriminadas as metas a atingir com a investigação e por fim, no tópico das opções metodológicas são caracterizadas as metodologias utilizadas para o desenvolvimento da investigação.

1.1. Problema de investigação, enquadramento e pertinência

A literacia financeira é cada vez mais uma temática importante para o dia-a-dia da sociedade, para além de pessoas formadas é também fundamental que estas tenham conhecimentos e capacidades de gerir os seus ativos e desta forma potenciar o seu futuro económico. Adicionalmente, as empresas também necessitam de potenciar os seus ativos e aumentar o seu *cash-flow*, de forma que estas consigam prosperar no mercado mais facilmente (Couto, 2013; Teles, 2011).

Existem diversas formas de gerar rendimentos, como por exemplo através de investimentos no imobiliário e no mercado de ações. Na ótica do imobiliário é possível rentabilizar o dinheiro investido através do arrendamento de imóveis ou através de compra e venda dos mesmos. Esta forma de investir caracteriza-se como um dos mais populares e mais comuns investimentos realizados pela população em geral para rentabilizar os seus ativos. Por outro lado, também é possível gerar um rendimento passivo e usufruir dos juros compostos através do mercado de ações. As ações são resultantes da tentativa por parte das empresas de aumentar o *cash-flow*, sem recorrer aos convencionais empréstimos bancários. Com a aquisição de uma ação de uma determinada empresa o investidor torna-se assim um acionista da organização, podendo participar de forma ativa na tomada de decisão da estratégia de futuro. Contudo, o sucesso ou insucesso da empresa vai fazer variar o valor da ação e desta forma o investidor vai estar exposto ao comportamento da mesma (Couto, 2013; Investor.gov, 2021; Teles, 2011).

Posto isto, é imperativo que a sociedade aumente o seu conhecimento acerca da gestão de ativo de modo a aumentar a rentabilidade dos seus ativos e possivelmente com recurso a investimentos no imobiliário e no mercado de ações. Além destes mercados também existem as *commodities*, que estão associados a metais metálicos como o ouro e o gás, ou, por outro lado, a materiais agrícolas como o trigo e o milho. Adicionalmente também existem as criptomoedas que estão associadas a moedas digitais e que têm ganho cada vez mais protagonismo e podem ser uma solução de futuro. De realçar que estes investimentos têm um risco diferente e consequentemente também têm um retorno diferente, contudo os investidores podem se expor a estes mercados caso se identifiquem com o seu perfil de risco. Atualmente é possível definir uma carteira de ativos com exposição em todos os mercados anteriormente e assim aumentar a diversificação dos investimentos, o que pode resultar em um crescimento sustentável dos mesmos (Chkir & Cosset, 2001; Frankenfield, 2022; Hayes, 2021a).

Assim, através deste projeto de investigação é pretendido a partir de um modelo de avaliação de ativos financeiros (CAPM) estimar as rentabilidades esperadas e o risco com base no comportamento histórico dos ativos e na sua análise aprofundada com recurso à aplicação de *Machine Learning*, com vista à constituição de uma carteira de investimentos eficiente, isto é, que

permita a maximização da rentabilidade esperada e a minimização do risco. Adicionalmente, é pretendido com este estudo o desenvolvimento de um modelo para a definição de uma carteira capaz de integrar diferentes tipos de ativos, com diferentes riscos associados e possibilitar a opção de um perfil de risco que vai resultar em uma carteira mais conservadora ou mais especulativa.

Este projeto de investigação poderá revelar-se uma mais-valia para a generalidade da população para potenciar e efetuar uma boa gestão de ativos financeiros sem necessitar de um grande conhecimento em investimentos. Para além disso, também poderá ser útil para as indústrias de gestão de patrimónios, fundos de pensões, bancos, fundos de investimento, entre outros.

Por último, para o desenvolvimento deste projeto será utilizada a metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM) e a Teoria do Portfólio como base científica. Relativamente ao software será utilizado um software de *low-code* denominado por *Konstanz Information Miner* (Knime) para a desenvolvimento do modelo de previsão de carteiras eficientes e por fim o método AHP e o método de divisão igualitária dos ativos de forma a encontrar qual o modelo que vai mais ao encontro dos objetivos definidos para o projeto.

1.2. Questão e objetivos de investigação

A questão que se pretende ver respondida com este projeto de investigação é: A otimização de uma carteira de ativos pode beneficiar da utilização de Aprendizagem Máquina (*Machine Learning*), maximizando a rentabilidade e minimizando o risco?

Posto isto, pretende-se desenvolver um processo capaz de calcular o valor intrínseco dos ativos através de um modelo de avaliação de ativos financeiros (CAPM), com base na análise efetuada, calcular as percentagens de exposição aos ativos disponibilizados de forma a alcançar a carteira eficiente que esteja associada a um maior retorno e por sua vez a um menor risco. Para isso vai ser necessário calcular o retorno e o risco associado aos ativos com base na Teoria de Portfólio.

Para além disso, com a execução deste projeto é pretendido alcançar, para um nível de risco moderado uma carteira de ativos que com base no histórico de dados até 2020 consiga superiorizar a rentabilidade do fundo de investimento com as 500 melhores empresas dos Estados Unidos da América, também designado por *Standard and Poor's 500* (S&P500).

Assim sendo, tendo em vista alcançar os objetivos pré-estabelecidos para o projeto espera-se que seja possível adquirir um conjunto de dados com qualidade suficientemente boa para não comprometer o sucesso do mesmo, preparar os dados eliminando todos os dados desnecessários ou com erros, integrar no modelo de CRISP-DM o processo de cálculo de CAPM, definir o modelo de previsão que apresenta melhores resultados para o cálculo do CAPM dos ativos no futuro, realizar testes de forma a analisar se foram alcançados os objetivos definidos inicialmente e, por último, organizar a informação de modo a que esta se torne de fácil interpretação para o público-alvo.

1.3. Opções metodológicas

Para o desenvolvimento do projeto de investigação “*Machine Learning* aplicado à Teoria do Portefólio”, relativamente às metodologias vai ser adotada uma perspectiva quantitativa visto que análise que será efetuada é com base num conjunto de dados e serão aplicadas modelos e teorias científicas. Segundo Coutinho (2011) esta perspectiva está alicerçada num modelo hipotético-dedutivo, pois o investigador assume que o problema de investigação tem uma solução objetiva, que pode ser encontrada recorrendo ao método científico, partindo da teoria. Desta forma, esta investigação tem como ponto de partida uma questão de investigação e o objetivo é testar, verificar e comprovar teorias e hipóteses. Assim, seguindo esta perspectiva é expectável que exista um estudo inicial do estado de arte que pode permitir definir objetivos a alcançar e hipóteses a verificar.

Depois será utilizada uma abordagem dedutiva pois será desenvolvido um estudo com base numa teoria e comprovada a sua fiabilidade para alcançar as carteiras eficientes através da maximização do retorno esperado e da minimização do risco. Algumas características destacadas por Machado & Davim (2020) sobre a abordagem em causa: “recolha de dados quantitativos; operacionalização de conceitos, para que possam ser medidos quantitativamente; abordagem muito estruturada; investigador independente do fenómeno que está a ser estudado; generalização baseada em amostras de grande dimensão”.

A nível da natureza da investigação esta será do tipo analítica ou explicativa de modo que o problema ou a situação é decomposto nos fatores que o influenciam, sendo estes analisados exhaustivamente de modo a avaliar o grau de influência dos mesmos em torno do objetivo do estudo que consiste nas carteiras eficientes (Machado & Davim, 2020; Oliveira, 2011).

Por último relativamente ao método de investigação será utilizado o experimental pois serão efetuadas comparações entre os valores reais e previstos para o mesmo período homólogo. Assim, vão ser definidos dois grupos, sendo um de controlo e outro experimental (Coutinho, 2011).

2. REVISÃO BIBLIOGRÁFICA

No capítulo da revisão bibliográfica é explicitada toda a pesquisa relevante que foi realizada no âmbito do *Machine Learning* e da Teoria de Portfólio. Desta forma, inicialmente são abordados temas relacionados com o *Machine Learning*, neste caso mais especificamente o modelo que vai ser utilizado para desenvolver o projeto de investigação, o CRISP-DM e além disso é também realizada um abreve apresentação do *Software Knime*. Posteriormente, são abordados alguns temas no âmbito financeiro como a rentabilidade e risco da carteira, na qual são explorados alguns exemplos práticos e são demonstradas as fórmulas de cálculo dos dois parâmetros. Depois a Teoria de Portfólio na qual se insere toda a metodologia que vai ser necessária aplicar para o correto desenvolvimento do projeto e por último uma breve análise aos tipos de investimento financeiros que são atualmente muito usados por investidores de forma a rentabilizar os seus ativos e que podem ser futuramente integrados no processo que vai ser desenvolvido de obtenção de carteiras eficientes.

2.1. CRISP-DM

Segundo Ferreira (2021) o *Data Mining* é a utilização de técnicas eficientes de análise de grandes bases de dados e de extração de úteis e possíveis inesperados padrões. Desta forma, as empresas que se focam na análise destes padrões nas bases de dados internas podem potenciar a empresa no mercado pois conseguem aprender mais sobre os seus clientes e consequentemente idealizar estratégias de marketing mais eficientes. Posto isto, visto que o *Data Mining* exige algum conhecimento e experiência na área, foram idealizados modelos que facilitam a sua aplicação para pessoas com menos experiência na Análise de Dados, sendo que um dos modelos o CRISP-DM.

O CRISP-DM é uma metodologia popular para aumentar o sucesso dos projetos de *Data Mining*. O modelo atual do processo de extração de dados fornece uma visão geral do ciclo de vida de um projeto de *Data Mining* (Chapman, 1999). Através da figura 1 são descritas de uma forma gráfica as diferentes fases associadas à metodologia CRISP-DM.

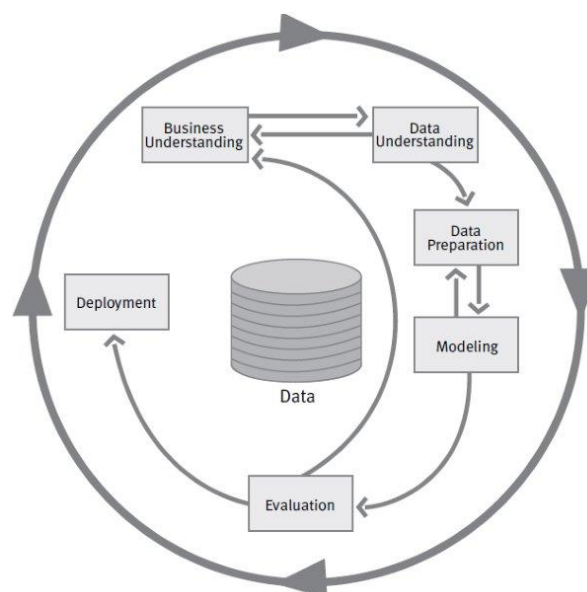


Figura 1 – Modelo CRISP-DM (Colpani, 2019)

O CRISP-DM define um projeto como um processo cíclico, em que várias iterações podem ser utilizadas para permitir um resultado final mais sintonizados com os objetivos pretendidos (Moro, Laureano, & Cortez, 2011). Nos seguintes subtópicos são apresentadas as 6 diferentes fases que integram esta metodologia, relativamente à sua importância e à sua execução.

2.1.1. Business Understanding

Esta fase inicial tem em foco a análise e compreensão dos objetivos do projeto e os requisitos numa ótica de negócio. De forma, a que posteriormente, com base no conhecimento alcançado, converter este projeto num problema de *Data Mining* e desenvolver um plano preliminar, tendo em vista alcançar os objetivos pré-estabelecidos (Chapman, 1999; Ferreira, 2021a).

Com base em Ferreira (2021a), o *Business Understanding* pode subdividir-se nos seguintes pontos:

- Determinar os objetivos na ótica do projeto;
- Avaliar situação atual;
- Determinar os objetivos de *data mining*;
- Desenvolver um plano inicial de projeto.

2.1.2. Data Understanding

Após definir os principais objetivos e de analisar o projeto é necessário proceder à análise da base de dados disponibilizada. Nesta fase, denominada por *Data Understanding*, é pretendido que seja efetuada uma análise ao *data set* e às variáveis que integram o mesmo. Além disso, é expectável que seja possível identificar possíveis problemas de qualidade de dados, interpretar os dados ou detetar subconjuntos interessantes para formar hipóteses da informação oculta (Chapman, 1999; Ferreira, 2021b; Moro, 2011). Posto isto, esta fase pode ser subdivida nos seguintes tópicos:

- Recolha de conjunto de dados iniciais;
- Descrição dos dados, relativamente ao tamanho da amostra e ao tipo de variáveis associadas ao *data set*;
- Explorar os dados, com recurso a gráficos como histogramas, gráficos de barras, gráficos de linhas, diagramas de pareto, *scatter plot*, *boxplots*, entre outros;
- Verificação da qualidade dos dados, por exemplo se os dados se encontram completos ou se não existem erros no *data set*.

2.1.3. Data Preparation

A fase de preparação dos dados tem como principal objetivo a construção de um conjunto de dados final a partir dos dados em bruto inicial. Esta fase também pode ser dividida em diferentes pontos, contudo não existe uma ordem em específico de execução dos pontos, ao invés das duas fases referidas anteriormente. Concretamente, para a preparação dos dados é necessário proceder à seleção, limpeza, construção, integração e formatação dos mesmos. Posto isto, é expectável que no final desta fase se obtenha um *data set* mais otimizado, objetivo, sem erros nos dados, com as

variáveis fulcrais para a realização do projeto e com os dados formatados corretamente, tendo em vista o alcance dos objetivos pré-estabelecidos no *Business Understanding* (Chapman, 1999; Ferreira, 2021b).

2.1.4. Modeling

Depois será necessário selecionar qual o modelo que vai ser utilizado no processo de *Machine Learning*. Existem diversos modelos que podem ser aplicados no processo de extração dos dados, contudo apenas é pretendido o modelo que assegura uma maior fiabilidade dos resultados gerados. Existem dois tipos de modelos, os de aprendizagem supervisionada que se subdividem em classificação e regressão, sendo que na classificação são utilizadas variáveis discretas, ao invés da regressão que utiliza variáveis contínuas. E os de aprendizagem não supervisionada em que os próprios modelos encontram padrões ou agrupamentos no conjunto de dados não etiquetados/rotulados, na medida em que o principal objetivo destes modelos é encontrar estruturas de dados relacionados, agrupar os mesmos de acordo com as semelhanças e representar o conjunto de dados num formato comprimido (Ferreira, 2021a, 2021d).

Posto isto, existem diversos modelos que podem ser aplicados ao projeto de investigação, primeiramente existem os modelos *Distance-based* que se subdividem no *Nearest Neighbour*, no *K-Nearest Neighbour* e no *Case-Based Reasoning*. O *Nearest Neighbour* é uma técnica de *Data Mining* que permite prever um *output* desconhecido, com base no passado do conjunto de dados. Esta técnica é baseada na memória, ao invés de aprender com um modelo compacto e pode ser aplicado nos modelos de aprendizagem supervisionada quer de classificação como de regressão. Para uma boa aplicação desta técnica e para um aumento da sua precisão é necessário que normalizar e preparar o conjunto de dados destinado ao treino. Depois, a técnica *K-Nearest Neighbour* difere da anterior no sentido em que considera os *k* dos vizinhos mais próximos de um objeto no conjunto de dados de treino. As principais vantagens desta técnica é o simples o algoritmo de treino, as funções objetivo mais complexas podem ser descritas por simples locais aproximações e têm uma natureza incremental. Contudo também apresenta algumas desvantagens como ser afetado por atributos que se caracterizam pela sua redundância e irrelevância, dificuldade em classificar um novo objeto e a densidade diminui com a dimensionalidade. Por fim a técnica *Cased-Based Reasoning* resolve um novo problema com base em soluções de problema similares no passado, sendo que o processo desta técnica consiste em recuperar casos relevantes para a resolução de um determinado problema, de seguida mapear a solução a partir do um caso similar no passado, depois testar a solução e caso necessário rever e por último depois de a solução ter sido adaptada com sucesso ao problema alvo, armazenar a experiência de resolução como uma novo caso em memória, de forma a que este auxilie no futuro a resolução de novos problemas (Ferreira, 2021d).

Além destes também é possível adotar modelos probabilísticos como o *Bayesian Learning*, o *Naive Bayes* e o *Bayesian Networks*. A técnica designada por *Bayesian Learning* assume que a probabilidade de um determinado evento "A" resultar em um evento "B", depende da relação entre os eventos "A" e "B" e da probabilidade de "A" acontecer independentemente de "B". Em relação à técnica *Naive Bayes* esta caracteriza-se como um dos mais eficientes e eficazes algoritmos de *Machine Learning* e *Data Mining*. O seu desempenho competitivo é surpreendente, porque a suposição da independência condicional na qual se baseia, raramente é verdadeira em aplicações

do mundo real (Zhang, 2004). Depois a técnica de *Bayesian Networks* utiliza a independência condicional para alcançar um equilíbrio entre o número de parâmetros calculados e a representação das dependências entre as variáveis. Representa a distribuição conjunta de probabilidade de um conjunto de variáveis aleatórias (Ferreira, 2021d).

Depois existem os métodos *Search-based* para os quais podem ser aplicadas técnicas como o *Decision Trees* e o *Regression Trees*. O *Decision Trees* é utilizado quando se tem uma grande quantidade de dados e a necessidade de os classificar para encontrar repostas sobre alguns assuntos. Para tal, pode-se utilizar o conceito de árvore de decisão como modelo que mapeia as observações, tendo em consideração um atributo selecionado como ponto de partida. Além disso, a técnica de *Decision Trees* auxilia na determinação do atributo que irá desenvolver um melhor desempenho na procura das informações necessária (Carneiro Da Rocha, Timoteo De Sousa Junior, & Timóteo De Sousa Júnior, 2010). Depois relativamente à técnica de *Regression Trees* normalmente a função de custo é o erro quadrático, minimizado pela média de amostras em cada folha da árvore e esta assemelha-se às árvores de classificação, tendo em conta a diferença na função de perda (Ferreira, 2021d).

Relativamente aos modelos *Optimization-based* existem várias técnicas associadas, contudo vão ser abordados duas técnicas: a *Neural Networks* e a *Support Vector Machine*. A *Neural Networks* caracteriza-se como uma técnica que se baseia nos dez biliões de neurónios do cérebro humano, idealizando assim neurónios artificiais que são baseados nos neurónios biológicos. Cada neurónio recebe um ou mais *inputs* e uma função de ativação é aplicada nos *inputs*, que determinam a saída do neurónio – o nível de ativação. As vantagens da aplicabilidade desta técnica definem-se como a maior precisão de classificação mesmo com problemas complexos, processamento distribuído, robusto tratamento de exemplos mesmo que estes integrem dados com erros, lidam bem com atributos redundantes uma vez que os pesos associados a eles são normalmente menores e os resultados podem ser discretos ou um vetor de valores (discretos ou reais). Contudo na nível de desvantagens é difícil determinar a topologia da rede ótima para um problema, pode demorar demasiado tempo a encontrar uma solução adequada, difícil de usar o que exige um longo tempo de formação e não é fácil de interpretar os resultados objetivos pois não é fornecido um modelo ou explicações sobre os resultados (Ferreira, 2021d).

2.1.5. Evaluation

Nesta fase do projeto foi construído um modelo ou modelos que se caracterizam pela sua qualidade na ótica da análise de dados. Contudo, primeiramente é importante avaliar mais profundamente o modelo e rever as etapas executadas para construir o modelo, de modo a ter a certeza de que este atinge corretamente os objetivos do projeto. O principal objetivo é determinar se existe algum objetivo do projeto que não foi considerado tal como é expectável. De forma a efetuar esta avaliação podem ser utilizadas métricas, técnicas e pode ser feita a comparação e seleção. Relativamente às métricas pode ser aplicada a matriz de confusão, a matriz de custo e a curva de ROC, as quais podem ser classificadas pela precisão, pelo *recall*, pela especificidade e pelo F1-score. No final desta fase deverá ser possível alcançar resultados em virtude da utilização do *data mining* (Chapman, 1999; Ferreira, 2021c).

2.1.6. Deployment

Após a criação do modelo, de forma concluir o projeto é necessário avançar para a fase de *deployment*. Esta fase tem como principais objetivos a organização, documentação e apresentação dos resultados obtidos a partir do modelo, visto que os resultados têm de ser de fácil interpretação para o utilizador final do modelo. Esta fase pode implicar por vezes desenvolvimentos de maior complexidade, de modo a moldar aos interesses do projeto, contudo também pode ser apenas gerar um *report* que se caracteriza como uma das soluções mais fáceis de executar. Além disso, também é necessário delinear um plano de manutenção do modelo, de forma a que este continue o mais otimizado possível e que represente uma boa solução a longo prazo para o projeto (Chapman, 1999).

2.2. Knime

O *Konstanz Information Miner* (Knime) é um software *open source* de análise de dados com um grande conjunto de blocos de construção e de ferramentas. Pode ser utilizado desde o carregamento do conjunto de dados até ao relatório final ou para prever novos valores utilizando um modelo. O Knime está disponível em diferentes versões, contudo apenas a versão para *Desktop* se define como *open source* (Bakos, 2013; O'hagan & Kell, 2015).

É um ambiente computacional modular, o que permite uma flexível e simples montagem visual, análise de dados interativa e processamento de dados. Caracteriza-se como uma plataforma de análise preditiva e um poderoso integrador de dados, concretamente o *software* auxilia na documentação das etapas complexas realizadas no pré-processamento, análise estatística, modelação estatística e análise preditiva (Feltrin, 2015; O'hagan & Kell, 2015).

O Knime fornece um vasto leque de funcionalidades para a execução de diferentes processos associados ao *Machine Learning*. As ferramentas desenvolvidas no *software* podem ser para uso privado ou publicado como contribuição comunitária. Sendo esta uma das principais características do Knime, a existência de um fórum *web* que auxilia na execução de muitos processos no *software*. O fórum também se encontra compatível com o programa o que possibilita que os utilizadores utilizem o *drag and drop* para colocar processos do site no *workflow* em desenvolvimento. Desta forma, o Knime fornece uma interface *user-friendly* que permite uma análise passo a passo dos *flows* executados (Barsacchi, 2013). Na figura 2 é apresentado a interface do *workflow* do Knime.

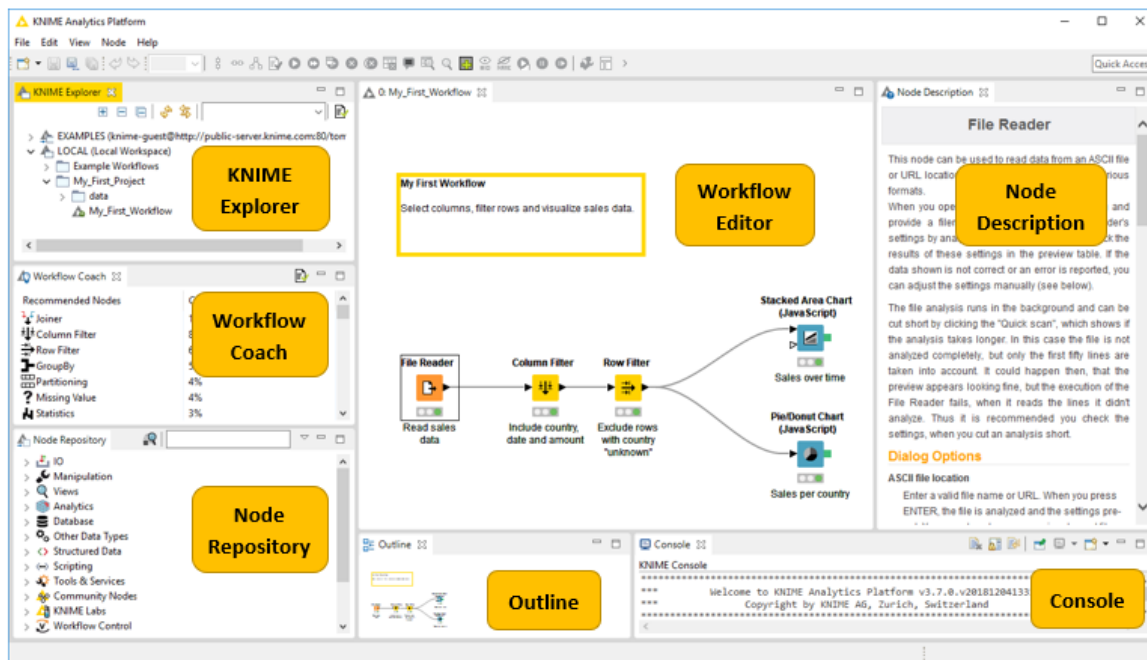


Figura 2 – Interface do software Knime (Knime, 2022)

2.3. Rentabilidade e Risco das Carteiras

A partir da análise de um estudo efetuado a três carteiras de valores mobiliários norte-americanos, mais concretamente, uma carteira de ações de empresas norte-americanas, uma carteira de obrigações de longo prazo do governo norte-americano e uma carteira de letras do tesouro, também designado por títulos da dívida do governo norte-americano, com vencimento a menos de um ano, foi possível perceber a rentabilidade e o risco associado às mesmas num conjunto de 109 anos. Foi definido como pressupostos que seria investido 1 dólar no ano de 1900 e o revestimento de todos os dividendos e juros pagos, de forma a efetuar uma comparação de rentabilidade entre as carteiras no ano de 2008. Deste modo, verificou-se que a carteira de ações se definia pela maior rentabilidade e, por outro lado, a carteira associada às letras do tesouro se caracterizava pela menor rentabilidade. Na tabela 1 são exibidos os valores das diferentes carteiras numa ótica nominal, isto é, sem ter em conta a inflação dos anos, e numa ótica real, onde é deduzido o valor associado à inflação nos 109 anos em estudo (Dimnson, 2021).

Tabela 1 – Valor resultante do investimento

| Valor resultante do investimento | | |
|----------------------------------|----------|--------|
| Carteira | Nominal | Real |
| Letras do Tesouro | \$71 | \$2,88 |
| Obrigações do Governo | \$242 | \$9,85 |
| Ações | \$14.276 | \$582 |

Consoante os valores obtidos é possível determinar a taxa média de retorno anual e o valor do prémio de risco, que se define como a taxa média de retorno anual nominal das ações menos a taxa

média de retorno nominal das letras do tesouro, conforme indicado na tabela 2. De realçar que as letras do tesouro estão associadas a um prémio de risco de 0%, pois estas acompanham a evolução natural do mercado.

Tabela 2 – Taxa média de retorno anual

| Taxa média de retorno anual | | | |
|------------------------------|---------|------|-----------------|
| Carteira | Nominal | Real | Prémio de risco |
| Letras do Tesouro | 4% | 1,1% | 0% |
| Obrigações do Governo | 5,5% | 2,6% | 1,5% |
| Ações | 11,1% | 8% | 7,1% |

Posto isto, é possível concluir que os investidores que optaram pela carteira de ações ganharam um prémio de risco de 7,1% sobre o retorno das letras do tesouro. Por outro lado, seria expectável que fosse associada à taxa de retorno do mercado (r_m) o valor de 11,1%, contudo esta não é a forma correta de proceder, visto que é improvável que o valor de r_m permaneça constante ao longo do tempo. Assim, os 7,1% definem-se como o valor mais sustentado, devido à sua elevada amostra de 109 anos, para a taxa de retorno do mercado. Deste modo, o procedimento mais correto para estimar o retorno expectável para o futuro é somar o valor do prémio de risco normal a taxa de juro sem risco (r_f), ou seja, a percentagem de rentabilidade futura das letras do tesouro. Em suma, é vital a precisão do valor associado ao prémio de risco de mercado para se efetuarem boas previsões para o futuro, contudo mesmo com mais de cem anos de dados em análise, é impossível comprovar a sua exatidão (Brealey, Myers, & Franklin, 2020; Dimnson, 2021).

Posteriormente, é necessário medir o risco da carteira, onde vão ser utilizadas medidas estatísticas normais de dispersão como o desvio padrão e a variância. Para efetuar o cálculo da variância, quando é estimada com base em uma amostra de retornos observados é utilizada a seguinte fórmula:

$$Variância = \frac{1}{N - 1} \sum (\bar{r}_m - r_m)^2$$

N – Número de observações

\bar{r}_m – Média do retorno efetivo do mercado no período

r_m – Retorno esperado do mercado no período

Para o exemplo em análise, das três carteiras distintas, procedeu-se ao cálculo dos valores da variância e desvio padrão, tendo em vista uma correta análise da variabilidade e risco associados a estes investimentos. Isto é mostrado na tabela 3 e permite concluir que a carteira de letras do tesouro tem um menor risco associado e carteira de ações uma maior variabilidade e consequentemente uma maior incerteza para os investidores.

Tabela 3 – Variabilidade das carteiras

| Carteira | Desvio padrão (σ) | Variância (σ^2) |
|-----------------------|----------------------------|--------------------------|
| Letras do Tesouro | 2,8 | 7,7 |
| Obrigações do Governo | 8,3 | 69,3 |
| Ações | 20,2 | 406,4 |

Depois, foram analisadas algumas ações em específico para perceber como diferentes valores de desvio-padrão, associadas às diferentes ações se iam comportar quando são integradas na mesma carteira. Após a análise efetuada foi possível concluir que a diversificação de ações reduz a variabilidade da carteira, isto porque os preços das ações não variam exatamente da mesma maneira. Na figura 3 é descrito o gráfico que relaciona o risco da carteira (desvio padrão) com o número de ações e que permitiu constatar a importância da diversificação das carteiras (Brealey, 2020; Henri, 2018).

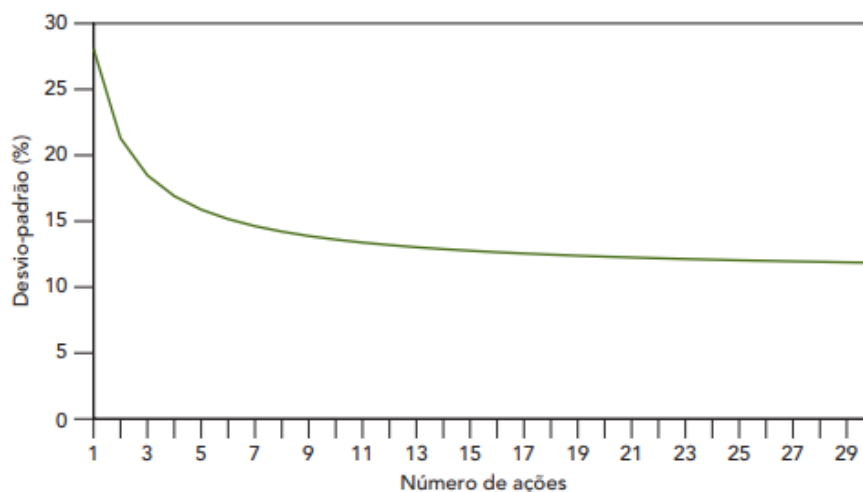


Figura 3 – Evolução do risco da carteira relativamente ao nº de ações (Brealey, 2020)

O risco que pode ser potencialmente eliminado por via da diversificação é designado por risco específico, está associado aos perigos específicos de cada empresa e também se define como risco não sistemático, risco residual, risco único ou risco diversificável. Contudo, conforme a figura 3 o gráfico tende a manter-se com um valor fixo de desvio padrão de aproximadamente 12%, este risco caracteriza-se pelo risco de mercado, este não pode ser evitado através da diversificação do portfólio pois é uma consequência da existência de outros riscos relativos a toda a economia e que afetam todos os negócios e também se denomina por risco sistemático ou risco não diversificável. O risco de mercado é o que leva à subida/descida da generalidade dos valores de ações. Na figura 4 é efetuada a legenda dos riscos que integram o gráfico de análise do desvio padrão relativamente à diversificação da carteira (Brealey, 2020; Henri, 2018).

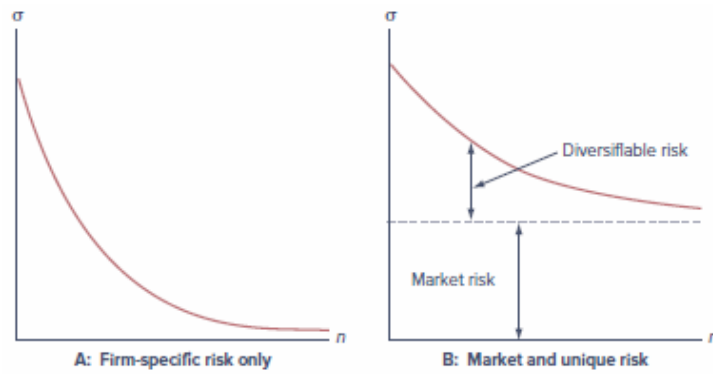


Figura 4 – Riscos da carteira (Henri, 2018)

Relativamente ao risco total da carteira, que pode ser minimizado com o aumento do número de ações, este deve ser calculado tal como a figura 5 destaca. Nesta figura apenas é efetuado o esquema para o cálculo da variância de 2 ações.

| | Ações 1 | Ações 2 |
|---------|--|--|
| Ações 1 | $x_1^2 \sigma_1^2$ | $x_1 x_2 \sigma_{12}$ $= x_1 x_2 \rho_{12} \sigma_1 \sigma_2$ |
| Ações 2 | $x_1 x_2 \sigma_{12}$ $= x_1 x_2 \rho_{12} \sigma_1 \sigma_2$ | $x_2^2 \sigma_2^2$ |

Figura 5 – Variância de uma carteira com duas ações (Brealey, 2020)

O procedimento para calcular a variância associada às duas ações consiste na soma dos diferentes valores dos diferentes blocos definidos. De uma forma mais simplificada, esta consiste na fórmula de determinação da variância para duas ações em uma carteira:

$$Variância = x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2 + 2(x_1 x_2 \rho_{12} \sigma_1 \sigma_2) \tag{1}$$

- ρ_{12} – Correlação entre as ações 1 e 2
- x_1 – Valor da ação 1
- x_2 – Valor da ação 2

Caso a carteira seja definida por mais do que duas ações, deve ser utilizada a seguinte fórmula de cálculo da variância:

$$Variância = \frac{1}{N} \times \text{variância média} + \left(1 - \frac{1}{N}\right) \times \text{covariância média} \tag{2}$$

N – Blocos de variância, ou seja, número de ações na carteira

Conforme é possível verificar através da fórmula definida, à medida que o N aumenta, a variância da carteira aproxima-se consistentemente da covariância média.

Posto isto, além da variância da carteira que pode ser minimizada pelo aumento das ações que integram a mesma, também é necessário considerar o risco de mercado das ações individualmente. Isto porque o risco de uma carteira bem diversificada depende do risco de mercado dos títulos incluídos na carteira. Para calcular este risco de mercado mencionado necessitamos de quantificar a sensibilidade em relação aos movimentos do mercado, esta que se define por beta (β). As ações com um beta superior a 1 tendem a ampliar o movimento geral do mercado, por exemplo no caso de uma ação com um beta de 1,41, caso o mercado suba 2% isto significa que a ação sobe 2 vezes o valor do beta, que resulta num aumento 2,82% da ação. Contudo, o inverso também acontece caso o mercado desça 2%. Depois, considerando o desvio padrão do mercado de 20%, caso uma carteira possua uma grande diversidade e quantidade de ações, o valor do beta tende para o 1, que se define como a correlação com o mercado. Assim, com um beta de 1,5 estaria associado um risco de carteira de 30% e, por outro lado, com um beta de 0,5 o risco da carteira definia-se nos 10%. Na figura 6 é demonstrado o gráfico de variação do risco da carteira pelo número de títulos, contudo com linhas associadas a diferentes valores de beta de forma a facilitar a interpretação do efeito do valor do beta na carteira.

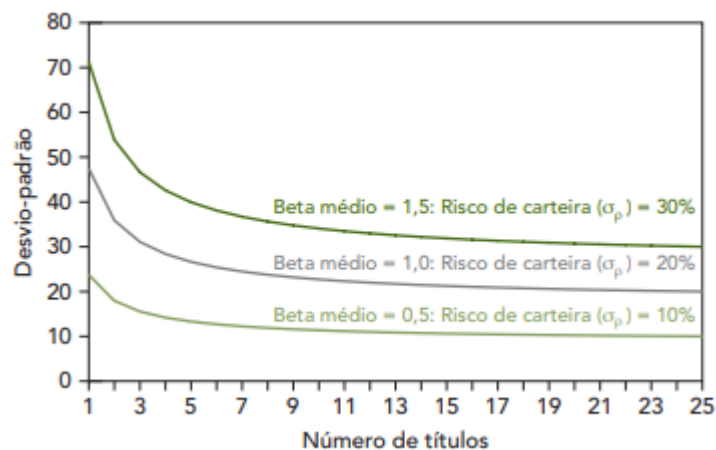


Figura 6 – Efeito da variação do beta nas carteiras (Brealey, 2020)

Assim, é possível concluir que o risco de uma carteira diversificada é proporcional ao beta da carteira, que é igual à média dos betas das ações que integram a mesma. Posto isto, tendo em vista o cálculo do valor do beta para os títulos da carteira deve-se ter em conta a seguinte fórmula:

$$\beta_i = \sigma_{im} / \sigma_m^2 \quad (3)$$

i – Ação

σ_{im} – Covariância entre os retornos da ação e do mercado

σ_m^2 – Variância dos retornos do mercado

Em suma, de forma a calcular o risco associado a uma determinada carteira de títulos, devemos recorrer ao cálculo da variância e do desvio padrão. De uma forma, geral uma carteira constituída

por ações tem uma maior rentabilidade e um maior risco associado, devido à sua variabilidade. Além disso, através da análise do estudo em causa verificou-se que a diversificação das ações se traduz na minimização do risco específico, contudo não tem efeito no risco do mercado que rege a generalidade das ações no mercado. Posto isto, percebeu-se que o risco de uma carteira bem diversificada depende do risco de mercado dos títulos incluídos na mesma, ou seja, no cálculo do risco e da rentabilidade da carteira é necessário integrar a variável beta associada à medição da sensibilidade em relação aos movimentos do mercado. A partir da análise do beta, constatou-se que o mesmo pode ter uma relação de proporcionalidade com o risco da carteira (Brealey, 2020; Dimnson, 2021).

De uma forma mais concreta o risco está associado à variabilidade das ações e outra forma de mitigar a sua variabilidade é através do *dollar-cost averaging* (DCA) que se caracteriza por uma estratégia de investimento em compras periódicas do mesmo valor numa determinada carteira de ações, tendo em vista a redução do impacto da volatilidade do mercado e consequentemente do risco associado a esta carteira (Hayes, 2021b).

2.4. Teoria do Portfólio

Com base no que foi analisado no capítulo anterior, foi possível constatar a importância do cálculo do desvio padrão para a determinação do risco das carteiras. Assim, o gráfico da distribuição normal da carteira de investimento deve ser similar ao definido na figura 7, ou seja, uma distribuição mais comprimida, onde estão associados menores valores de desvio padrão e com um enviesamento para os valores positivos. Caso uma distribuição se caracterize por uma maior variabilidade, esta se define mais estendida e consequentemente está associado a uma carteira de investimento com um maior risco associado (Bodie, Kane, & Marcus, 2013; Brealey, 2020; Elton, Gruber, Brown, & Goetzmann, 2013; Fabozzi & Markowitz, 2011).

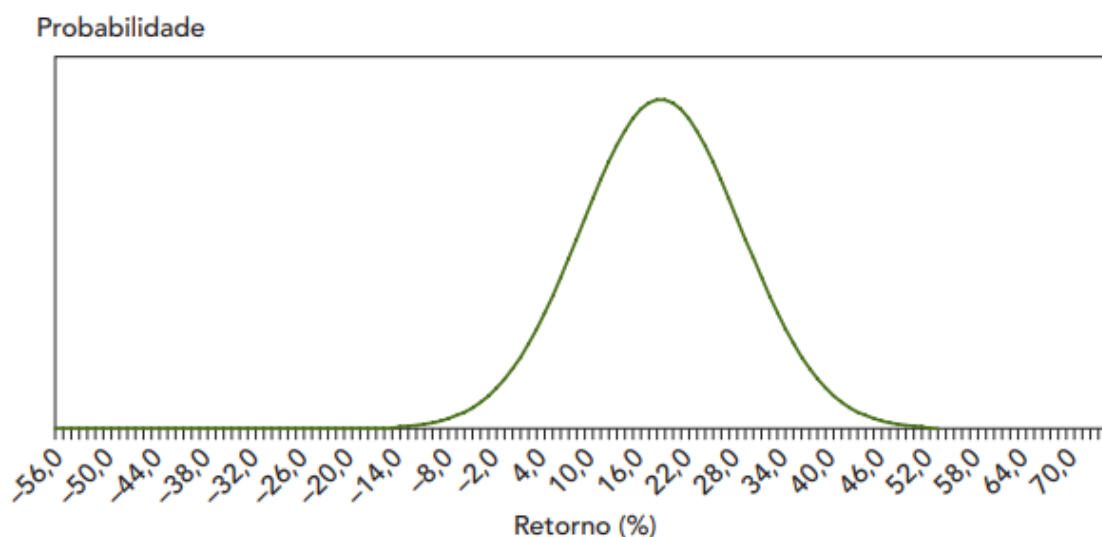


Figura 7 – Distribuição Normal de Investimento (Brealey, 2020)

Contudo, quando uma carteira possui várias ações isto significa que temos diferentes gráficos de distribuição normal entre as diversas ações, contudo, tendo em conta o retorno esperado de cada ação e o seu desvio padrão é possível analisar como deve ser diversificada a carteira em termos de percentagem de aquisição de cada ação. Posto isso, utilizando os valores de retorno esperado e

desvio padrão de 10 ações foram definidas 3 diferentes carteiras que se diversificam com percentagens de alocação de investimento distintas, tal como é possível ver na tabela 4.

Tabela 4 – Carteiras eficientes escolhidas de dez ações

| Ação | Retorno Esperado | Desvio padrão | Percentagens de Carteiras | | |
|-------------------------------------|------------------|---------------|---------------------------|--------------|-------------|
| | | | A | B | C |
| Amazon | 22,8% | 50,9% | 100% | 10,9% | - |
| Ford | 19% | 47,2% | - | 11% | - |
| Dell | 13,4% | 30,9% | - | 10,3% | - |
| Starbucks | 9% | 30,3% | - | 10,7% | 3,6% |
| Boeing | 9,5% | 23,7% | - | 10,5% | - |
| Disney | 7,7% | 19,6% | - | 11,2% | - |
| Newmont | 7% | 36,1% | - | 9,9% | 10,2% |
| Exxon Mobil | 4,7% | 19,1% | - | 9,7% | 18,4% |
| Johnson & Johnson | 3,8% | 12,5% | - | 7,4% | 33,9% |
| Campbell Soup | 3,1% | 15,8% | - | 8,4% | 33,9% |
| Retorno esperado da carteira | | | 22,8% | 10,5% | 4,2% |
| Desvio padrão da carteira | | | 50,9% | 16% | 8,8% |

A partir da tabela acima descrita, no gráfico da figura 8 é possível perceber o retorno esperado e o desvio padrão referentes às ações descritas referente às carteiras definidas, estas que se caracterizam sob a forma de um losango e cruz respetivamente. Além disso, a área sombreada representa as combinações possíveis entre os retornos esperados e os desvios padrão. Visto que as carteiras A, B e C estão posicionadas ao longo da linha com maior espessura, estas definem-se como carteiras eficientes. Em suma, o objetivo é perceber qual a carteira sob a linha delimitadora da carteira que se caracteriza por um maior retorno esperado e um menor desvio padrão (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

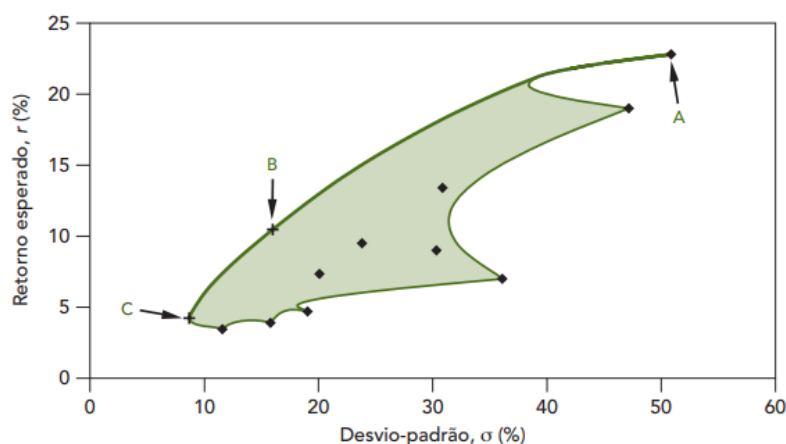


Figura 8 – Gráfico de Rentabilidade e Risco de Carteiras (Brealey, 2020)

De forma a determinar a melhor carteira eficiente, devemos começar pelo ponto r_f no eixo vertical e traçar a linha mais inclinada possível para a curva das carteiras eficientes, de forma que seja tangente à curva, tal como apresentado na figura 9. Posto isto, a carteira eficiente no ponto de tangência designa-se pela melhor em relação às outras possibilidades. O índice que relaciona o retorno esperado do investimento e o desvio padrão é designado por Índice de *Sharpe* e pode ser calculado através da seguinte fórmula:

$$\text{Índice de Sharpe} = \frac{\text{Prémio de risco}}{\text{Desvio padrão}} = \frac{r - r_f}{\sigma} \quad (4)$$

σ – Desvio-padrão

r – Retorno esperado

r_f – Retorno esperado das letras de tesouro

Este índice é muito utilizado pelos investidores para medir o desempenho ajustado ao risco de gestores de investimento. Além disso, na ótica do investidor um aumento do valor investido por meio de um empréstimo a uma taxa sem risco vai aumentar as suas possibilidades de retorno, contudo também o desvio padrão associado. Por outro lado, caso o investidor empreste uma parte do dinheiro investido na carteira, a sua posição no mercado vai ser sob a reta entre os pontos r_f e S (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

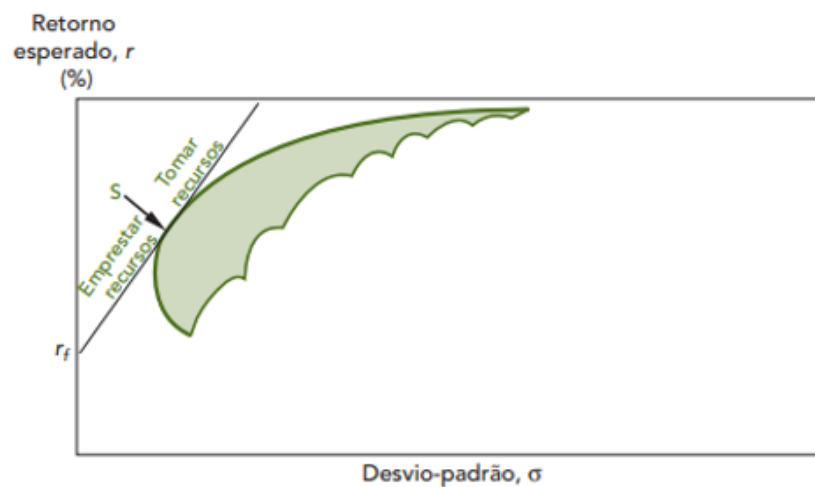


Figura 9 – Efeito dos empréstimos na carteira (Brealey, 2020)

Posto isto, um investidor deve ter em conta dois principais aspetos, em primeiro lugar é necessário escolher a melhor carteira de ações, que no exemplo apresentado se define como a letra S. Posteriormente, é fulcral decidir a combinação da carteira com emprestar ou pedir emprestado para se obter uma exposição ao risco do agrado do investidor (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

Em meados da década de 1960, três economistas – William Sharpe, John Lintner e Jack Treynor – desenvolveram um modelo de avaliação de ativos de capital, ou CAPM, que auxilia na determinação do prêmio de risco, tendo em conta o valor do beta. O modelo CAPM estabelece que o prêmio de risco esperado de cada investimento é proporcional ao seu beta. Conforme apresentado na figura 10 a linha de mercado de títulos tem um declive variável devido ao prêmio de risco ($r_m - r_f$), assim é possível concluir que os investimentos deviam de ficar situados ao longo da linha inclinada.

Adicionalmente, visto que existe uma relação de proporcionalidade entre o beta e o prémio de risco, para um beta de 0,5 é, portanto, estimado metade do prémio de risco esperado do mercado. Contudo, caso o beta tenha o valor de 2 é estimado o dobro do valor do prémio de risco esperado do mercado. Esta relação é expressa a partir da seguinte forma:

Prémio de risco esperado das ações = beta × prémio de risco esperado do mercado

$$(\Rightarrow) r - r_f = \beta(r_m - r_f)$$

(5)

r – Retorno esperado

r_f – Retorno esperado das letras de tesouro

$(r_m - r_f)$ – Prémio de Risco

β – Sensibilidade de movimentos no mercado

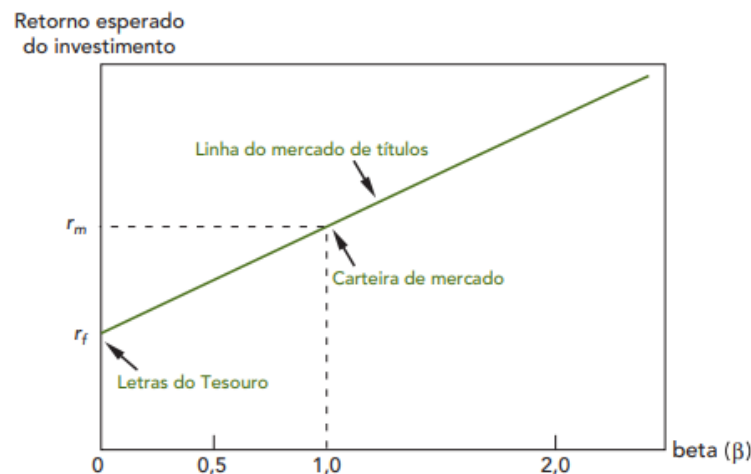


Figura 10 – Gráfico descritivo do modelo CAPM (Brealey, 2020)

A média da generalidade das ações que constituem a carteira de mercado terá uma posição sob a linha de mercado de títulos, contudo caso seja feita uma análise individualizada será possível perceber algumas ações abaixo a linha e, por outro lado, outras ações acima da linha. Caso nenhuma se situe abaixo da linha, então também não poderá haver nenhuma acima da linha. O objetivo dos investidores não consiste no investimento em ações que se encontrem abaixo da linha de mercado de títulos, pois podem obter resultados melhores se fizerem um investimento mais diversificado e, conseqüentemente, na linha em causa. Assim, os investidores procuram maximizar o seu retorno com o menor beta possível, sendo assim natural que invistam em empresas que eles acreditem que se encontra subvalorizadas e que podem levar a retornos superiores ao mercado (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

Uma grande maioria dos gestores financeiros acreditam que o CAPM se define como a ferramenta mais conveniente para lidarem com o cálculo do risco dos investimentos e para estimar o custo do capital. Contudo o CAPM ainda apresenta algumas ineficiências, mais concretamente foi feita uma descoberta de que o CAPM não tem em conta no seu cálculo a dimensão das empresas e o índice valor contábil-valor de mercado. Desta forma, alguns investigadores acreditam que uma versão simples do modelo pode não representar toda a verdade. Além disso, um segundo problema

associado a este modelo é que a carteira de mercado deveria conter todos os investimentos com risco, incluindo ações, obrigações, mercadorias, imóveis e até mesmo capital humano, contudo neste momento a grande maioria dos índices de mercado contém apenas uma amostra de ações (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

Atualmente, existem outros modelos alternativos ao CAPM como a teoria da avaliação por arbitragem (APT) e o modelo dos três fatores. A APT, de Stephen Ross, não pergunta a eficiência das carteiras, em vez disso, parte do princípio de que o retorno de cada ação depende, parcialmente, de influências macroeconômicas pouco claras em parte de ruído, que se define como acontecimentos que são específicos das empresas na sua individualidade. Nesta teoria não é utilizada o valor da carteira do mercado, ao contrário do CAPM, e permite-nos moldar o cálculo ao retorno aos fatores que o investidor entender que sejam importantes. Depois, relativamente ao modelo dos três fatores de Fama e French (1992), o retorno esperado de cada uma das ações depende da sua exposição a cada um dos fatores sugeridos: o retorno da carteira de mercado menos a taxa de juro livre de risco, a diferença entre o retorno das ações de pequenas e grandes empresas e a diferença entre o retorno das ações com índices mais elevados e reduzidos de valor contábil-valor de mercado. Assim é possível concluir que este modelo, mitiga um dos problemas associados ao CAPM relativamente à dimensão das empresas e ao índice valor contábil-valor de mercado, o que permite assim complementar o cálculo do CAPM (Bodie, 2013; Brealey, 2020; Elton, 2013; Fabozzi & Markowitz, 2011).

2.5. Tipos de Investimentos Financeiros

Neste capítulo são discriminados os diferentes tipos de investimento que podem ser utilizados para a idealização de uma carteira de ativos. Estes são analisados individualmente relativamente à sua rentabilidade e ao seu risco.

2.5.1. Ações

As ações consistem em constituintes de uma empresa, sendo que um investidor com a compra de uma ação detém uma percentagem da empresa onde se encontra locado o seu investimento. Assim, o investidor é designado por acionista e pode ser parte das decisões de futuro da empresa. Esta é uma solução utilizada por várias organizações, de forma evitar contrair um novo empréstimo ao banco. Normalmente as ações oscilam com base nos lucros ou perdas da empresa, contudo também se caracterizam como um investimento especulativo. Além disso, como na generalidade dos produtos no mercado o preço das ações também varia consoante a oferta e a procura da mesma. Existem duas formas de um investidor lucrar com uma ação, sendo que em primeiro lugar um investidor pode lucrar com a compra da ação por um valor inferior e posterior venda da mesma a um preço mais elevado. Por outro lado, um investidor também pode lucrar com os dividendos resultantes da posse da ação, este cenário apenas se aplica caso as empresas distribua os seus lucros na forma de dividendos. Além disso, de realçar que o investimento em ações pode ser lucrativo para os investidores, contudo também pode ser arriscada. Para efetuar um correto investimento nas ações é necessário ter um grande conhecimento sobre o valor das empresas e ter capacidade de prever oportunidades de mercado (Degiro, 2022).

2.5.2. Commodities

As *commodities* caracterizam-se por um mercado de matérias-primas ou produtos primários, que podem ser divididas em duas grandes categorias, as duras e as leves. As duras designam-se por recursos naturais que podem ser minerados ou extraídos como o exemplo do ouro, borracha, petróleo, entre outros. Por outro lado, as leves estão normalmente associadas a produtos agrícolas ou pecuários como o milho, o trigo, o café, o açúcar, a soja e a carne suína. Certas *commodities*, como os metais preciosos, são considerados uma boa proteção contra a inflação, sendo que vários investidores também confiam nas *commodities* durante os períodos de maior volatilidade no mercado, visto o preço das mesmas tende a mover-se no sentido oposto das ações. Desta forma, seria uma boa estratégia adquirir uma maior percentagem de *commodities* na carteira quando se prevê uma quebra no mercado. Assim sendo, este investimento está associado a um menor risco e a uma menor rentabilidade em relação às ações (Hayes, 2021a).

2.5.3. Real Estate

O *Real Estate* está associado ao Imobiliário e consiste num dos investimentos mais utilizados pela globalidade da sociedade. Existem cinco categorias principais de imóveis, entre elas a residencial, a comercial, a industrial, o terreno e a finalidade especial. No quotidiano é frequente associar-se o investimento em imobiliário na compra ou no aluguer de um imóvel. As principais vantagens associadas a este investimento consistem na sua rentabilidade estável, na valorização do capital investido e na diversificação do portfólio. Por outro lado, o imobiliário está normalmente associado a um investimento ilíquido, é influenciado por fatores locais, requer um grande investimento inicial e pode exigir alguma experiência para a gestão dos mesmos. Contudo também é possível investir em imóveis indiretamente, sendo que uma das formas mais populares de o fazer é através de *Real Estate Investment Trust* (REIT). Estes fundos caracterizam-se por empresas que detêm um portfólio de imóveis que possuem uma renda. Através da aquisição de REITs é possível tornar os investimentos em *Real Estate* mais líquidos e transparentes, além disso não é necessário que os investidores possuam uma grande experiência em gestão de imóveis e, adicionalmente, também é prescindível um grande investimento de capital na compra do imóvel. Por outro lado, os REITs também possuem alguns contras, entre eles a baixa valorização do capital investido, a inexistência de vantagens fiscais, a exposição ao risco de mercado e as taxas altas (Chen, 2021).

2.5.4. Criptomoedas

Uma criptomoeda caracteriza-se como uma moeda digital ou virtual protegida por criptografia, que impossibilita a sua falsificação ou o pagamento com a mesma múltiplas vezes. Muitas destas moedas utilizam a tecnologia da blockchain que consiste num banco de dados que é compartilhado entre os nós de uma rede de computadores, de forma a manter um registo seguro e descentralizado de transações. Além disso, como as criptomoedas não são emitidas por nenhuma autoridade central, estas tornam-se imunes à interferência ou manipulação do governo. O Investimento em criptomoedas trata-se do que tem um maior risco associado devido a volatilidade destes mercados, contudo como se trata de um dos principais investimentos de futuro, estas devem ser analisadas para integrar no futuro uma carteira de ativos. Atualmente a moeda mais popular e valiosa é a *Bitcoin* e pode ser adquirida através do investimento na mesma e conseqüente aquisição de bitcoin,

como também através da sua extração. Além da Bitcoin, outra moeda que se distingue pela sua popularidade é a *Ethereum*. De uma forma geral, as criptomoedas têm todas uma função e especificação que as difere e, deste modo, algumas criptomoedas podem se adequar mais ao perfil do investidor ou invés de outras (Frankenfield, 2022; Hayes, 2022).

2.6. Método AHP

O método *Analytic Hierarchy Process* (AHP) foi desenvolvido por Tomas L. Saaty no início da década de 70 tendo como principal objetivo auxiliar na tomada de decisão na seleção de alternativas num problema com múltiplos critério/atributos. Caracteriza-se por um método de avaliação hierárquica de atributos com possibilidade de ser aplicado às diversas áreas de estudo por possibilitar que sejam desenvolvidas análise qualitativas e subjetivas através de cálculos matemáticos. Assim, com esta metodologia pretende-se aproximar ao máximo da decisão mais correta com base nas alternativas existentes e nos seus atributos (Ben & Ufrgs, 2006; Marins, Souza, & Barros, 2009; Trevizano & Freitas, 2005).

Segundo (Costa, 2002), para a construção e utilização de um modelo de estabelecimento de prioridades fundamentado no uso da metodologia AHP é necessário que sejam desenvolvidas as seguintes etapas:

- **Construção de hierarquias:** problemas complexos podem ser mais bem interpretados e compreendidos através da utilização de níveis hierárquicos. Assim, é necessário que no primeiro nível seja definido o propósito ou o objetivo do sistema, depois os critérios ou atributos que vão ser utilizados para classificar as diferentes alternativas e, por conseguinte, no último nível hierárquico encontram-se definidas as diferentes alternativas. Na figura 11 é apresentada uma estrutura exemplo dos níveis hierárquicos que são necessários construir para a aplicação da metodologia AHP (Costa, 2002; Marins et al., 2009; Trevizano & Freitas, 2005).

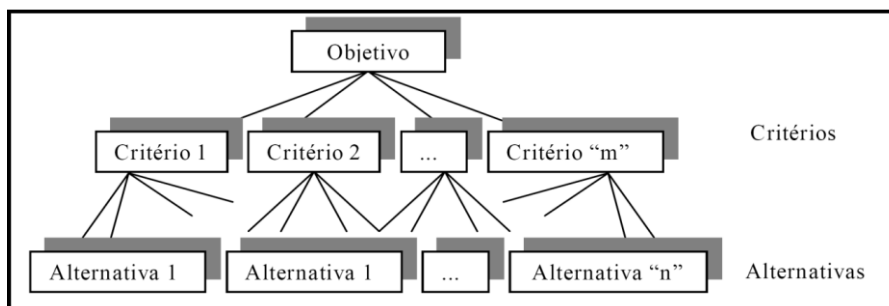


Figura 11 - Estrutura Hierárquica de Método AHP

- **Definição de prioridades:** esta etapa consiste no desenvolvimento de matrizes entre os níveis hierárquicos, de forma que todas as alternativas possuam um valor associado a cada critério em utilização. Após o desenvolvimento das diferentes matrizes para cada critério é necessário efetuar a sua normalização através da divisão de cada elemento destas matrizes pelo somatório dos valores da respetiva coluna. Posteriormente é necessário efetuar a o cálculo da média dos valores calculados anteriormente através da divisão pelo número total de alternativas. Por último, com base na significância definida para cada critério é efetuado o cálculo de um vetor de prioridades global, que armazene a prioridade associada

a cada alternativa em relação ao foco principal (Costa, 2002; Marins et al., 2009; Trevizano & Freitas, 2005).

- **Consistência lógica:** que se caracteriza pela avaliação do modelo de priorização construído relativamente à sua consistência (Costa, 2002; Marins et al., 2009; Trevizano & Freitas, 2005).

3. CASO DE ESTUDO

Neste capítulo será explicitado o estudo efetuado que envolve ferramentas como o Knime onde são efetuadas as análises das variáveis do *dataset* e as previsões utilizadas no projeto e como o Excel no qual foi desenvolvido uma metodologia AHP que auxilia na determinação da carteira de ativos. Numa primeira é descrita toda a análise efetuada a nível de interpretação dos dados, depois relativamente aos modelos de previsão dos ativos e por fim é discriminada a ferramenta utilizada para otimizar os investimentos tendo em conta os objetivos pretendidos.

3.1. Business Understanding

O objetivo principal consiste no desenvolvimento de modelos que aumentam o conhecimento relativamente ao tema da Teoria do Portefólio e auxilie a tomada de decisão de entidades ou da sociedade na sua generalidade. Os modelos em causa terão como foco a capacidade de previsão da variação dos valores dos diferentes ativos disponíveis no mercado através dos seus dados históricos. De forma que os modelos integrem toda a variedade de ativos disponíveis no mercado, procedeu-se à compilação dos diversos ativos no mesmo período de análise. Um dos principais objetivos é a realização de uma análise de como os diferentes ativos podem estar interrelacionados. Posto isto, as análises efetuadas serão utilizadas para a definição de carteiras de ativos que estejam associadas a uma maior taxa de rentabilidade e a uma maior diversificação, o que se traduz num menor risco de aquisição das mesmas.

Assim sendo, os principais objetivos do projeto assentam nos seguintes pontos:

- Prever para o período pretendido, com base no histórico dos dados, o valor de mercado dos ativos financeiros;
- Prever para o período pretendido, com base noutras variáveis, o valor de mercado dos ativos financeiros;
- Prever para o período pretendido, qual a constituição da carteira de ativos que maximiza os lucros e minimiza o risco do investimento;
- Verificar relações de proporcionalidade entre os diferentes ativos em análise;
- Constatar quais os ativos que estão associados a uma maior volatilidade;
- Verificar se os valores dos ativos oscilam positivamente ou negativamente consoante o período do ano em que se encontram.

Posto isto, numa fase inicial do projeto será realizado um pré-processamento dos dados com o objetivo de remover informação que não seja pertinente e de normalização dos mesmos. Posteriormente, é necessário proceder ao processamento dos dados que tem como principal objetivo retirar informação sob a forma de *gráficos*, *boxplots*, análises de correlação e outras ferramentas que permitam a constatação de aspetos relevantes para a realização deste projeto. De forma que numa próxima fase seja possível determinar quais os modelos que otimizam a previsão dos diferentes ativos que integram o conjunto de dados.

Por último, será efetuada uma análise comparativa entre o método AHP e o método de divisão igualitária dos ativos da carteira, tendo em vista percepção do modelo que potencia os resultados do projeto.

3.2. Data Understanding

Nesta fase do projeto procedeu-se à análise dos dados recolhidos, relativamente às relações entre os atributos do *dataset*. Assim neste tópico será abordada a forma de como foram recolhidos de dados, será efetuada uma discriminação das variáveis que constituem o *dataset*, serão explorados os dados com auxílio de elementos gráficos e por último será analisada a qualidade dos dados, tendo em conta os principais objetivos que se pretendem alcançar com o desenvolvimento deste projeto.

3.2.1. Recolha de Dados

O conjunto de dados obtido, tratam-se dados reais que retratam a evolução mensal dos ativos no mercado de ações. Inicialmente, foi efetuado o *download* de diferentes bases de dados de forma a perceber qual a informação que poderia ir mais ao encontro dos objetivos do projeto. Posto isto, de uma forma geral vão ser utilizados dados associados a diferentes *ETF's*, depois vai ser utilizado o índice *S&P500*, assim como variáveis que auxiliam na previsão deste índice e por fim também diversas *Commodities* que poderão potenciar a diversificação das carteiras de ativos resultante do projeto desenvolvido.

3.2.2. Descrição dos Dados

Neste ponto procedeu-se à descrição dos atributos que constituem o *dataset*. Assim nas tabelas 5, 6 e 7 são discriminadas os atributos relativamente ao seu significado e à sua tipologia. De realçar que para facilitar a sua análise os mesmos foram agrupados em 3 grandes grupos: *ETF's*, *S&P500* e *Commodities*. Sendo que o período de recolha de dados se definiu a partir do mês de janeiro de 2020 e foram apenas selecionados os valores associados ao dia um de cada mês.

Tabela 5a – Atributos do Grupo *ETF's*

| ETF's | | |
|----------|-------------------------------|---|
| Atributo | Tipologia | Descrição |
| ETF_DIA | Variável Numérica Contínua | Nome: <i>American Century Disciplined Growth Fund Y Class</i> Categoria: <i>Large Value</i> |
| ETF_EEM | Variável Numérica Contínua | Nome: <i>Baird Ultra Short Bond Fund Institutional Class</i> Categoria: <i>Diversified Emerging Mkts</i> |
| ETF_EWZ | Variável Numérica Contínua | Nome: <i>iShares Russell 1000 Large-Cap Index Fund Investor A Shares</i> Categoria: <i>Latin America Stock</i> |

Tabela 5b – Atributos do Grupo ETF's

| ETF's | | |
|----------|-------------------------------|---|
| Atributo | Tipologia | Descrição |
| ETF_ARKG | Variável Numérica Contínua | Name: American Century One Choice Blend+ 2040 Portfolio A Class Categoria: Health |
| ETF_FCG | Variável Numérica Contínua | Name: American Century Global Gold Fund R Class Categoria: Equity Energy |
| ETF_ASHR | Variável Numérica Contínua | Name: American Century One Choice Blend+ 2045 Portfolio Investor Class Categoria: China Region |
| ETF_BBRE | Variável Numérica Contínua | Name: Thrivent Municipal Bond Fund Class A Categoria: Real Estate |
| ETF_COPX | Variável Numérica Contínua | Name: American Century Global Gold Fund A Class Categoria: Natural Resources |
| ETF_EFG | Variável Numérica Contínua | Name: Invesco Balanced-Risk Commodity Strategy Fund Class C Categoria: Foreign Large Growth |

Tabela 6 – Atributos do Grupo S&P500

| S&P500 | | |
|-------------------|-------------------------------|--|
| Atributo | Tipologia | Descrição |
| S&P500 | Variável Numérica Contínua | Variação de mercado do Índice S&P500 |
| EUR_USD | Variável Numérica Contínua | Variação de mercado da Relação entre o dólar e o euro |
| GDP USA | Variável Numérica Contínua | Variação de mercado do Produto Interno Bruto dos Estados Unidos da América |
| Taxas de Juro USA | Variável Numérica Contínua | Variação de mercado das Taxas de Juro dos Estados Unidos da América |
| Gold | Variável Numérica Contínua | Variação de mercado do Ouro |

Tabela 7 – Atributos do Grupo Commodities

| Commodities | | |
|---------------------|----------------------------|---------------------------------------|
| Atributo | Tipologia | Descrição |
| Crude Oil Petroleum | Variável Numérica Contínua | Variação de mercado do Petróleo Bruto |
| Aluminum | Variável Numérica Contínua | Variação de mercado do Alumínio |
| Beef | Variável Numérica Contínua | Variação de mercado da Carne de Vaca |
| Coal | Variável Numérica Contínua | Variação de mercado do Carvão |
| Copper | Variável Numérica Contínua | Variação de mercado do Cobre |
| Cotton | Variável Numérica Contínua | Variação de mercado do Algodão |
| Rice | Variável Numérica Contínua | Variação de mercado do Arroz |
| Tea | Variável Numérica Contínua | Variação de mercado do Chá |
| Uranium | Variável Numérica Contínua | Variação de mercado do Urânio |
| Wheat | Variável Numérica Contínua | Variação de mercado do Trigo |

3.2.3. Exploração dos Dados

Neste tópico foram analisadas ao detalhe as relações entre as diversas variáveis que constituem o *dataset*. De forma que seja efetuada uma análise que vá mais ao encontro ao objetivo do projeto foram definidos 3 grupos de análise sendo eles os definidos de seguida:

- *ETF's*, que possui todos os *ETF's* que constituem a amostra de dados. Os ativos em causa subdividem-se nas seguintes categorias: Mercados diversificados emergentes, ações da américa latina, grande valor de mercado, região da china, saúde, energias limpas, recursos naturais, grande crescimento e imobiliário.
- *S&P500*, que integra os valores do índice *S&P500* assim como variáveis que podem relacionar-se com as oscilações do índice em causa, como o exemplo do ouro, o produto interno bruto e as taxas de juro dos Estados Unidos da América e o valor do dólar em relação ao euro;

- *Commodities*, que contemplam todos ativos que se definem como mercadorias primárias que servem como matéria-prima para o fornecimento de diferentes setores da sociedade como o exemplo do petróleo, alumínio, carne de vaca, carvão, cobre, algodão, arroz, chá, urânio e trigo.

Numa primeira fase foram analisados os atributos que integram o grupo do S&P500, para o qual dado a elevada quantidade de dados, procedeu-se à análise dos valores a partir do primeiro mês de 2000. Na figura 12 é definido o gráfico onde é possível perceber a variância dos valores relativos de cada ativo do grupo S&P500.

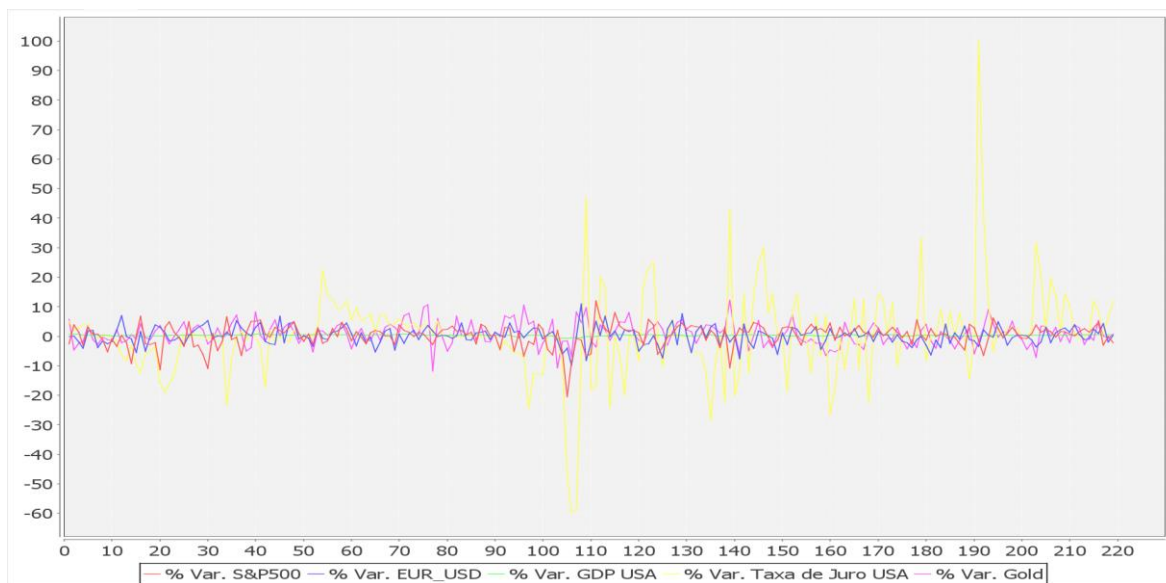


Figura 12 – Variação relativa dos ativos do grupo S&P500

Contudo, a partir do gráfico da figura 12 não é fácil verificar relações entre os diferentes ativos, posto isto procedeu-se à adição de colunas ao *dataset* onde é efetuado o cálculo da variação relativa acumulada ao longo do intervalo em análise. Assim, foi desenvolvido um novo gráfico à semelhança do demonstrado anteriormente.

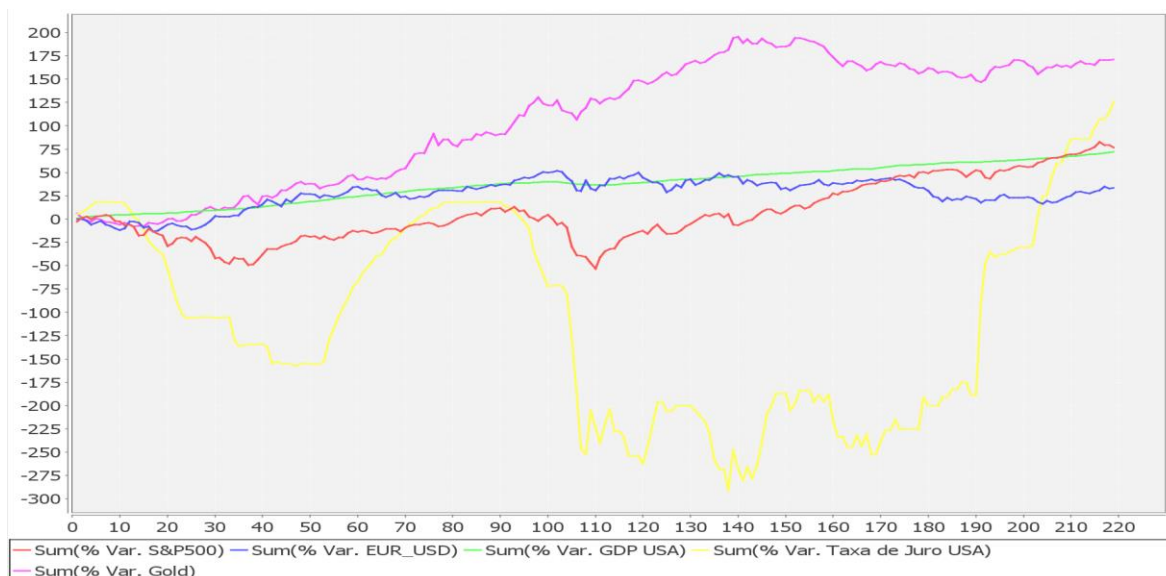


Figura 13 – Variação relativa acumulada dos ativos do grupo S&P500

Após a análise do gráfico foi possível verificar que o ouro, o S&P500 e a relação entre o euro e o dólar ao longo dos anos oscilam da mesma forma o que nos dá indícios da existência de uma correlação entre as mesmas. Em contrapartida as taxas de juro dos Estados Unidos da América vão contrariando parcialmente a variação dos outros ativos mencionados anteriormente, o que também pode resultar numa correlação entre os ativos, contudo de uma forma negativa. De realçar que os dados relativos ao produto interno bruto dos USA vão ao longo dos anos acompanhando a variação relativa dos outros ativos, mas com menores oscilações tratando-se assim de um atributo mais contante.

Assim, de forma a analisar a relação entre os diversos ativos procedeu-se ao desenvolvimento de uma matriz de correlação dos ativos, que se encontra definida na tabela 8 e na figura 13.

Tabela 8 – Correlação entre os ativos do grupo S&P500

| Atributo 1 | Atributo2 | Valor de Correlação |
|------------------------------|------------------------------|----------------------|
| Sum(% Var. S&P500) | Sum(% Var. EUR_USD) | 0.19992980753301032 |
| Sum(% Var. S&P500) | Sum(% Var. GDP USA) | 0.8144962378464778 |
| Sum(% Var. S&P500) | Sum(% Var. Taxa de Juro USA) | 0.23867653781732825 |
| Sum(% Var. S&P500) | Sum(% Var. Gold) | 0.621148967640516 |
| Sum(% Var. EUR_USD) | Sum(% Var. GDP USA) | 0.6200684123907705 |
| Sum(% Var. EUR_USD) | Sum(% Var. Taxa de Juro USA) | -0.40257641618647033 |
| Sum(% Var. EUR_USD) | Sum(% Var. Gold) | 0.7395501286481092 |
| Sum(% Var. GDP USA) | Sum(% Var. Taxa de Juro USA) | -0.11936257357760185 |
| Sum(% Var. GDP USA) | Sum(% Var. Gold) | 0.9215198931148366 |
| Sum(% Var. Taxa de Juro USA) | Sum(% Var. Gold) | -0.3802483548847236 |

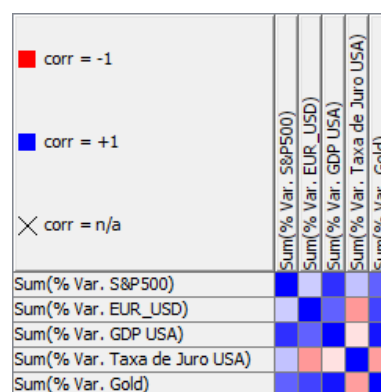


Figura 14 – Matriz de correlação dos ativos do grupo S&P500

A análise de correlação entre as variáveis permitiu confirmar a existência de uma correlação positiva entre as seguintes variáveis: o produto interno bruto dos EUA, o ouro, a relação entre o dólar e o euro e o S&P500. Em contrapartida também foi possível constatar a correlação negativa entre a Taxa de Juro dos EUA, a relação entre o dólar e o euro e o ouro. Assim, é possível perceber que com estas variáveis é possível realizar a previsão do S&P500 através das outras variáveis que

integram o grupo em causa. No apêndice A é apresentada a estrutura de nós que foi desenvolvida para estudar os atributos constituintes do grupo de S&P500.

Depois a nível das variáveis que constituem o grupo de *ETF's* foi também efetuada uma análise relativa dos seus valores desde o início de 2020 até aos valores mais atuais.

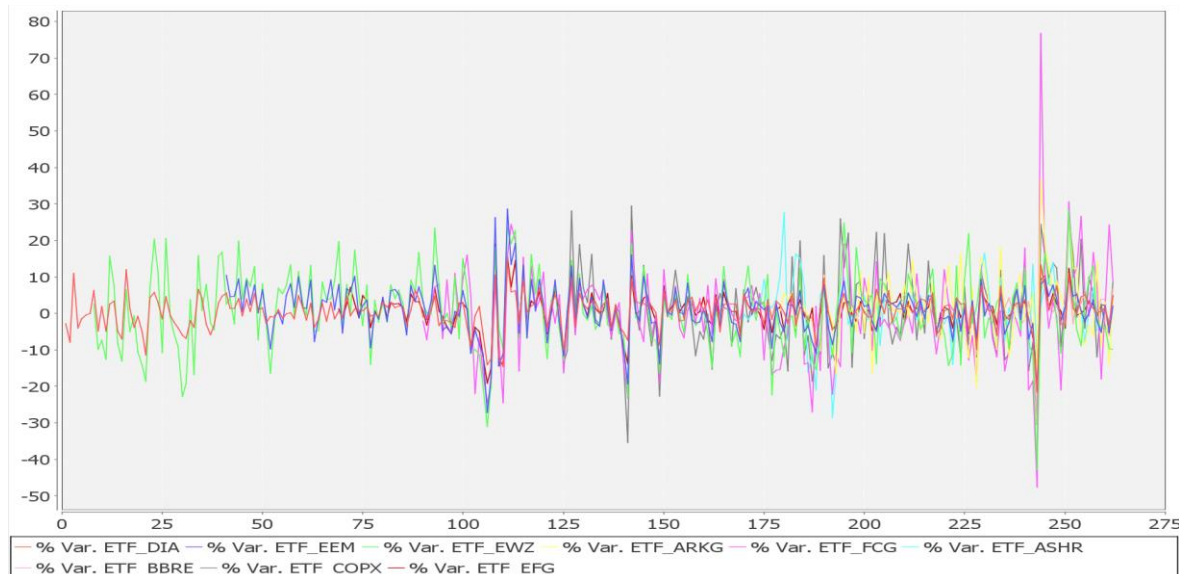


Figura 15 – Variação relativa dos ativos do grupo *ETF's*

Contudo à semelhança do acontecido anteriormente, esta análise também não nos permite alcançar o pretendido. Posto isto, procedeu-se ao desenvolvimento de um gráfico de comparação entre os atributos relativamente ao valor acumulado variação relativa dos mesmos. Assim sendo, na figura 16 encontra-se definido o gráfico referido.



Figura 16 – Variação relativa acumulada dos ativos do grupo *ETF's*

Desta forma, é possível verificar que os diferentes ativos ao longo dos anos grandes descidas e grandes subidas na mesma altura, o que pode estar associado a fatores externos como por exemplo a situação da pandemia vivenciada durante os últimos anos. Apesar disso, analisando a variação global dos ativos percebe-se que ativos mais constantes ao longo dos anos que outros, o que

leva a um menor risco de investimento nos mesmos. Além disso, também é possível destacar que existem ativos que por vezes têm uma relação inversa relativamente a outros ativos, o que pode realçar que se definem como atributos que combinamos podem potenciar a diversificação da carteira de ativos e conseqüentemente minimizar o risco associado.

Posto isto, foi efetuado uma análise das correlações entre os diferentes atributos de forma a destacar todas as relações entre os mesmos, visto que podem ser benéficas para a definição da carteira de ativos ideal, tendo em conta os objetivos de maximizar o lucro e minimizar o risco.

Tabela 9 – Maiores correlações entre os ativos do grupo ETF's

| Atributo 1 | Atributo2 | Valor de Correlação |
|----------------------|----------------------|---------------------|
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_EEM) | 0.750493580586525 |
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_ARKG) | 0.913041999314706 |
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_FCG) | -0.688956292035986 |
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_ASHR) | 0.702873610004599 |
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_BBRE) | 0.873713121141471 |
| Sum(% Var, ETF_DIA) | Sum(% Var, ETF_EFG) | 0.953957001523403 |
| Sum(% Var, ETF_EEM) | Sum(% Var, ETF_EWZ) | 0.920495866559841 |
| Sum(% Var, ETF_EEM) | Sum(% Var, ETF_ARKG) | 0.851375618483442 |
| Sum(% Var, ETF_EEM) | Sum(% Var, ETF_BBRE) | 0.750617977681286 |
| Sum(% Var, ETF_EEM) | Sum(% Var, ETF_COPX) | 0.790322780910587 |
| Sum(% Var, ETF_EEM) | Sum(% Var, ETF_EFG) | 0.836498279985938 |
| Sum(% Var, ETF_EWZ) | Sum(% Var, ETF_COPX) | 0.791454702147123 |
| Sum(% Var, ETF_ARKG) | Sum(% Var, ETF_ASHR) | 0.775950943077115 |
| Sum(% Var, ETF_ARKG) | Sum(% Var, ETF_COPX) | 0.81660819327986 |
| Sum(% Var, ETF_ARKG) | Sum(% Var, ETF_EFG) | 0.954898579219802 |
| Sum(% Var, ETF_FCG) | Sum(% Var, ETF_EFG) | -0.527692016151225 |
| Sum(% Var, ETF_ASHR) | Sum(% Var, ETF_EFG) | 0.754958572021824 |
| Sum(% Var, ETF_BBRE) | Sum(% Var, ETF_COPX) | 0.748142056539378 |
| Sum(% Var, ETF_BBRE) | Sum(% Var, ETF_EFG) | 0.801241223935439 |

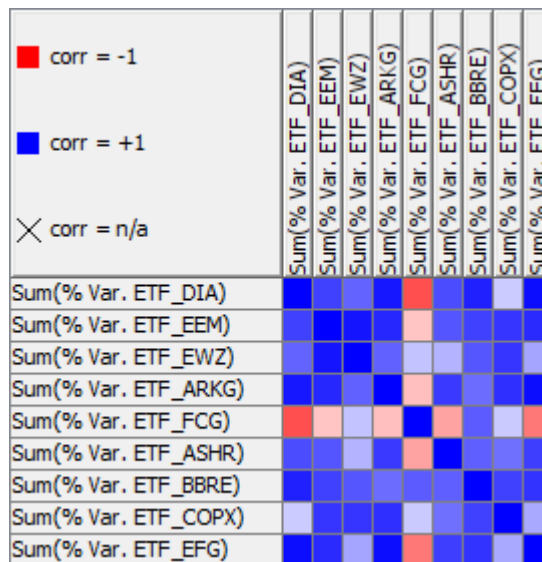


Figura 17 – Matriz de correlação dos ativos do grupo ETF's

A partir da análise das correlações entre as variáveis foi possível constatar que de um modo geral todos os *ETF's* selecionados possuem uma correlação positiva entre os mesmos, contudo existem também correlações negativas. De realçar a correlação negativa entre o ETF DIA e o ETF FCG e entre o ETF_EFG e o ETF_FCG. Contudo, o principal objetivo desta análise para o grupo dos *ETF's* era destacar quais os ativos que possuem uma menor correlação e que conseqüentemente levam a que uma combinação entre os mesmos esteja associado a uma carteira de ativos mais diversificado, isto é, com um menor risco.

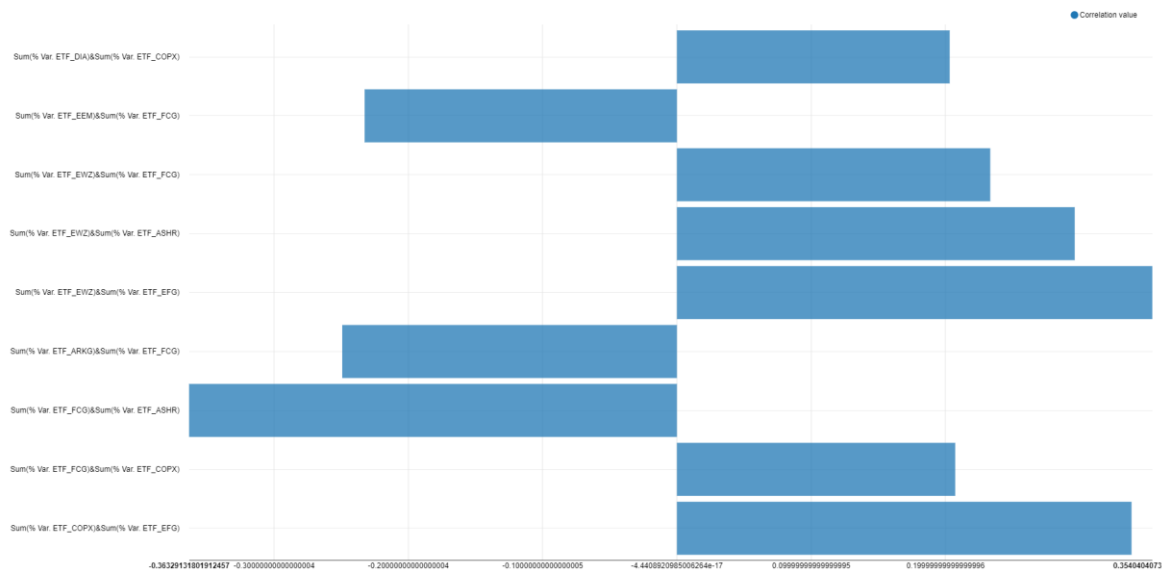


Figura 18 – Gráfico de menores correlações do grupo ETF's

Na figura 18 encontram-se destacadas as correlações entre atributos do *dataset* mais próximas de zero, que vão potencializar a diversificação da carteira de ativos caso sejam utilizadas as combinações em causa na definição do portefólio. Este elemento gráfico é importante visto que um dos objetivos do projeto é que a proposta de investimentos resultantes do processo se caracterize pela que possui menor risco, tendo em conta os outros objetivos definidos. Posto isto, são enumeradas de seguida, por ordem crescente, as combinações que foram alcançadas a partir da análise efetuada:

- ETF DIA e COPX;
- ETF FCG e COPX;
- ETF EEM e FCG;
- ETF EWZ e FCG;
- ETF ARKG e FCG;
- ETF EWZ e ASHR;
- ETF COPX e EFG;
- ETF EWZ e EFG;
- ETF FCG e ASHR.

No apêndice B é apresentado a estrutura de nós que foi desenvolvida para estudar os atributos constituintes do grupo de *ETF's*.

Por último para o grupo das *Commodities* foi desenvolvida uma análise á semelhança das desenvolvidas anteriormente. Tal como nos grupos anteriores, foram apenas filtrados os dados a partir de janeiro de 2000, de forma que todos os data sets se encontrem associado ao mesmo período. Assim, numa primeira fase foi efetuada criação de um elemento gráfico da variação relativa dos ativos do grupo em análise, contudo a sua análise também não se verificou fácil de alcançar alguma conclusão sobre as relações entre os dados.

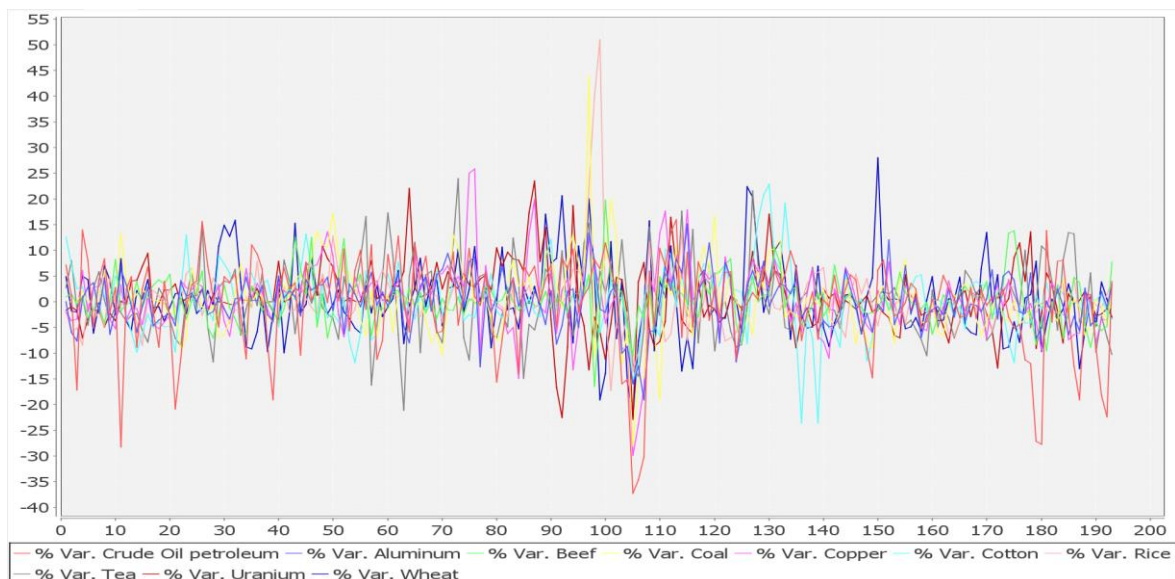


Figura 19 – Variação relativa dos ativos do grupo *Commodities*

Posto isto, procedeu-se então ao desenvolvimento do gráfico de linhas com a variação acumulativa dos atributos do grupo tendo como principal objetivo a identificação de relações entre os ativos. Deste modo, na figura 20 é possível analisar a variação relativa acumulada dos diferentes atributos.

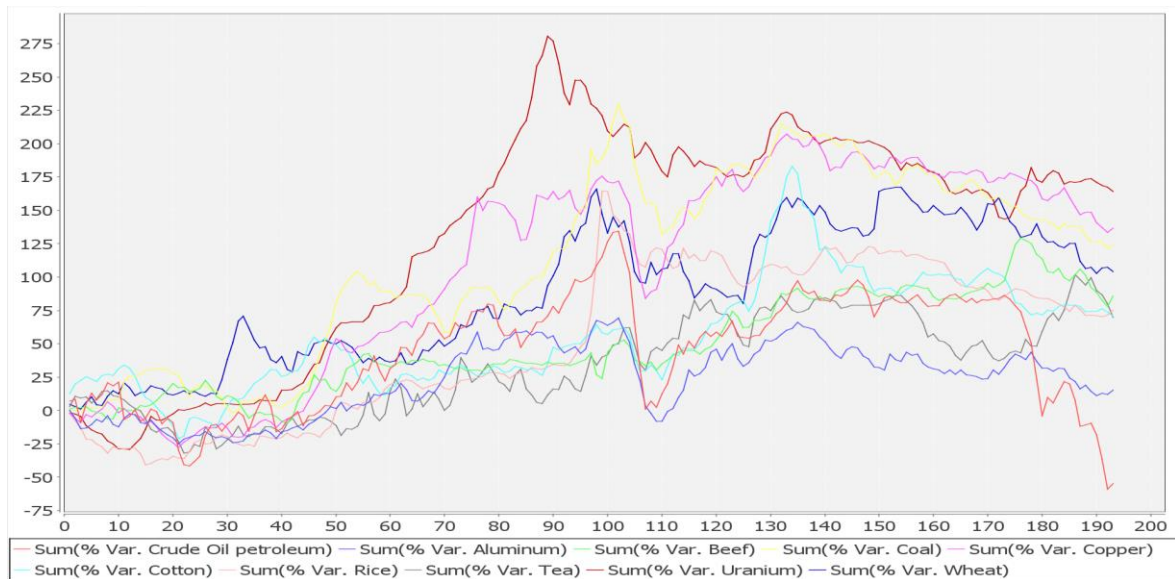


Figura 20 – Variação relativa acumulada dos ativos do grupo Commodities

Com base no gráfico definido acima é possível verificar que os atributos que compõem o grupo *Commodities* têm uma grande variabilidade e poucos se relacionam com outros atributos. Desta forma, a partir desta primeira análise constatamos que podem-se traduzir em ativos potenciadores da diversificação do portefólio. Posto isto, procedeu-se à análise das correlações entre os diferentes atributos que integram o grupo das *Commodities*.

Tabela 10 – Maiores correlações entre os ativos do grupo Commodities

| Atributo 1 | Atributo2 | Valor de Correlação |
|----------------------|---------------------|---------------------|
| Sum(% Var. Aluminum) | Sum(% Var. Copper) | 0.8959152568324878 |
| Sum(% Var. Aluminum) | Sum(% Var. Uranium) | 0.8803687274752906 |
| Sum(% Var. Coal) | Sum(% Var. Copper) | 0.925892771467231 |
| Sum(% Var. Coal) | Sum(% Var. Rice) | 0.9490670290904171 |
| Sum(% Var. Coal) | Sum(% Var. Uranium) | 0.867916949923861 |
| Sum(% Var. Coal) | Sum(% Var. Wheat) | 0.8972267613914706 |
| Sum(% Var. Copper) | Sum(% Var. Rice) | 0.8891211139043724 |
| Sum(% Var. Copper) | Sum(% Var. Uranium) | 0.9329253391355716 |
| Sum(% Var. Copper) | Sum(% Var. Wheat) | 0.9020511200479177 |
| Sum(% Var. Rice) | Sum(% Var. Tea) | 0.8772921134823863 |
| Sum(% Var. Rice) | Sum(% Var. Wheat) | 0.8843696197348017 |

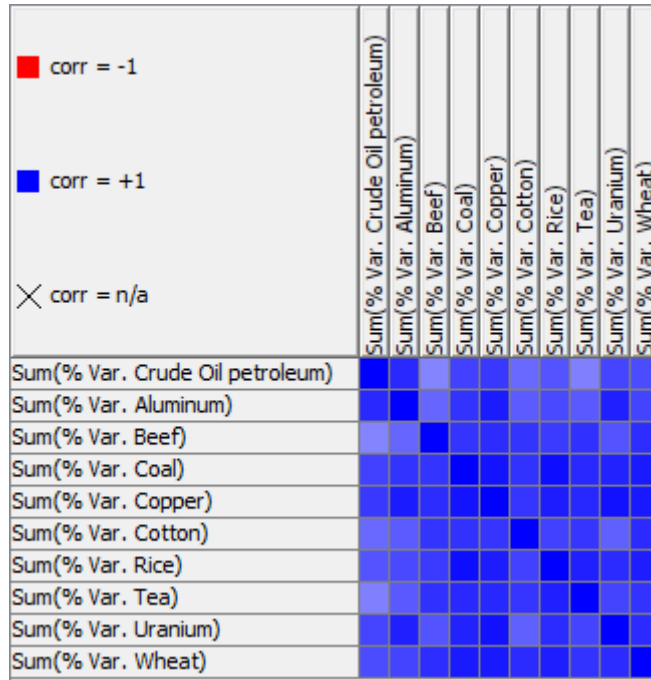


Figura 21 – Matriz de correlação dos ativos do grupo Commodities

Na tabela 10 estão definidas as maiores correlações entre as diferentes *Commodities* e na figura 21 é exibida matriz de correlação, sendo que a partir da análise destes dois elementos concluímos que existe uma elevada relação de proporcionalidade direta entre os diferentes atributos. Posto isto, estes ativos caracterizam-se pela sua simetria nível de oscilações do mercado o que consequentemente implica uma menor diversidade da cadeira de ativos. Contudo, um dos principais objetivos desta análise de dados também consiste na perceção das combinações de ativos que potenciam o projeto, neste caso através da diversificação.

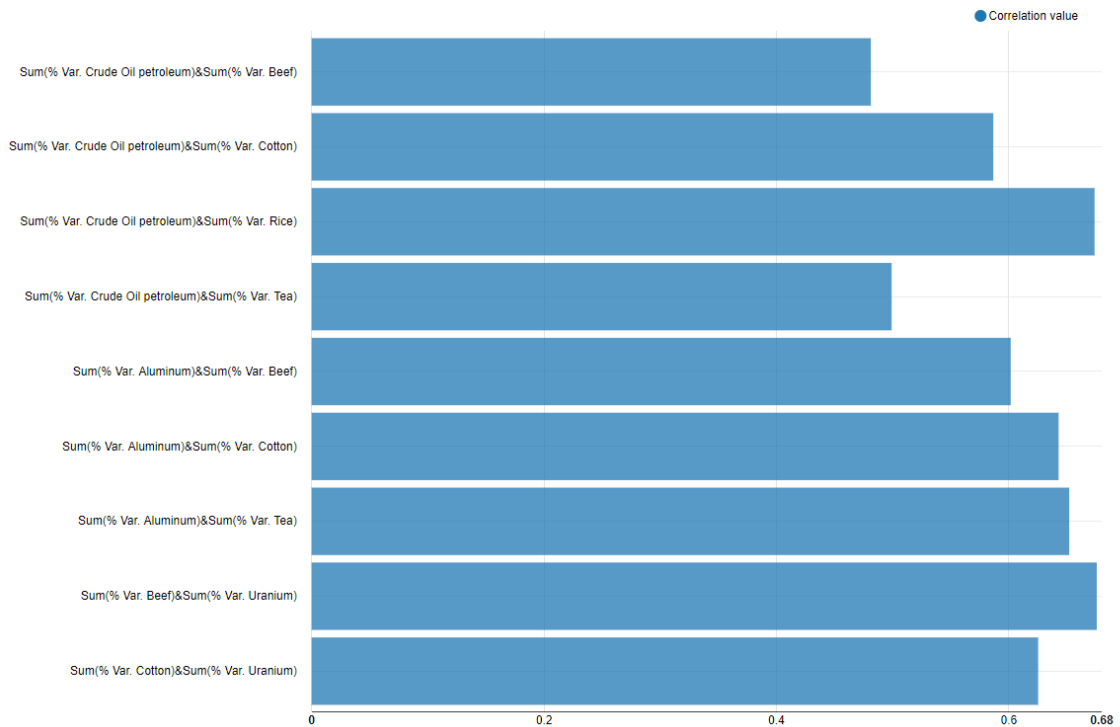


Figura 22 - Gráfico de menores correlações do grupo Commodities

Assim, na figura 22, encontram-se destacadas as principais combinações de ativos que apresentam uma correlação inferior e que podem potencializar a diversificação caso integrem a mesma carteira de ativos. Deste modo, já é possível verificar quais as combinações de ativos que podem ir mais ao encontro do pretendido, que consiste na maximização dos lucros do investimento e na minimização do risco associado. No apêndice C é apresentada a estrutura de nós que foi desenvolvida para estudar os atributos constituintes do grupo de *Commodities*.

3.3. Data Preparation

No decorrer da fase de *Data Understanding* verificou-se que existiam alguns atributos com erros e com escalar cronológicas desajustadas em relação ao pretendido. Assim, a nível de *Data Preparation* procedeu-se à seleção, limpeza e preparação dos dados para a futura etapa de modelação.

3.3.1. Seleção de Dados

Para o projeto em causa foram explorados diversos conjuntos de dados de ativos do mercado financeiro, contudo apenas se selecionaram os conjuntos de dados que tinham uma maior qualidade tendo em conta os objetivos pretendidos.

Deste modo, de forma que fosse possível definir uma carteira de ativos diversificada e com o menor risco possível foram selecionados ETF's de diferentes categorias e capazes de aumentar a flexibilidade das carteiras. Além dos ETF's também foram selecionados as *commodities* mais utilizadas no mercado financeiro. Com as *commodities* o objetivo é tornar as carteiras de ativos independentes da variabilidade do mercado empresarial e conseqüentemente aumentar a variabilidade de ativos disponíveis.

Depois, foi selecionado o índice S&P500 que se caracteriza como um dos índices mais populares atualmente, visto que integra as 500 melhores empresas dos EUA. Por último, também associado ao índice S&P500 foram destacadas as principais variáveis que poderiam auxiliar na previsão do mesmo, possibilitando assim numa fase posterior efetuar a comparação das previsões do S&P500 através do seu passado e através das variáveis selecionadas.

3.3.2. Limpeza e Preparação de Dados

Nesta fase de limpeza e preparação do knime procedeu-se à reposição dos valores em falta (através da utilização do *node "Missing Value"*) pelo valor 0 no grupo S&P500 visto que os mesmos se encontravam no início do *dataset*. Depois para o grupo dos *ETF's*, uma vez que existem valores em falta no meio do conjunto de dados, procedeu-se à reposição dos mesmos através de uma interpolação linear.

Além da reposição dos valores em falta também se efetuou a ordenação do conjunto de dados e à filtragem dos mesmos, de forma que os dados fiquem compreendidos entre janeiro de 2000 e os dados mais atuais.

De salientar que a nível de *Data Understanding* foram utilizados os *nodes "Lag Column"* e "*Math Formula*" que possibilitaram a definição de uma percentagem de variação entre os valores do mês

x em relação ao mês x-1. Posteriormente, através das colunas de variação mensal relativa foram desenvolvidas colunas de variação acumulada dos diferentes atributos através do *node* “*Moving Aggregation*”, o que permitiu a idealização da análise dos dados descrita anteriormente.

3.4. Modeling

A inteligência artificial é cada vez mais utilizada por empresas e organizações para prever determinados fatores e indicadores que auxiliam na hábil tomada de decisão por parte das mesmas. Deste modo, na fase de *Modeling* vão ser analisados os principais modelos preditivos tendo em conta o *dataset* em utilização e por fim será efetuada a previsão dos atributos constituintes do conjunto de dados para auxiliar na determinação da carteira de ativos ótima consoante os objetivos definidos.

3.4.1. Seleção das Técnicas de Modelação

Antes de realizarmos a previsão dos atributos é necessário identificar quais os modelos preditivos que se enquadram melhor com o conjunto de dados em utilização. Visto que todos os atributos constituintes dos 3 grupos do *dataset* se caracterizam pela sua tipologia numérica foram utilizados os seguintes modelos de previsão:

Simple Regression Tree – as árvores de regressão básicas dividem os dados em conjuntos mais pequenos e depois aplicam um modelo simples para cada subgrupo, de modo a determinar a relação entre diversas variáveis independentes e uma variável dependente;

Linear Regression – consiste numa equação que permite estimar o valor esperado de um atributo-alvo, através de um restante conjunto de variáveis. A relação é denominada linear, tendo em conta que se considera que a relação de resposta às variáveis é uma função linear de alguns parâmetros;

Polynomial Regression – consiste num tipo de regressão segundo a qual as variáveis independentes e a variável dependente serão modeladas por um polinómio de grau n;

Random Forest (Regression) – baseia-se num algoritmo de aprendizagem supervisionado que utiliza o método de ensemble *learning* para efetuar a regressão. Por sua vez, *ensemble learning* consiste numa técnica que combina previsões de múltiplos algoritmos de *machine learning* para realizar uma previsão mais precisa num só modelo. No caso de regressão, o output será a média dos valores obtidos das várias árvores;

Gradient Boosted Trees (Regression) – constitui métodos de ensemble que realizam regressão ou classificação através da combinação de outputs obtidos de árvores individuais. Utiliza a combinação de diversas *decision trees*, de modo a reduzir o risco de *overfitting*. O método de construção e resultados obtidos resulta do método de *boosting*, que combina “*weak learners*”, ou seja, *decision trees* com apenas uma divisão sequencialmente, para que as novas árvores corrijam os erros das anteriores;

Tree Ensemble (Regression) – o método efetua a aprendizagem de *ensemble trees*, utilizando metodologias semelhantes à *Random Forest*, contudo disponibiliza outras funcionalidades mais dinâmicas, como a utilização de divisões pela média para atributos numéricos ou a divisão binária para atributos nominais.

3.4.2. Modelo de Teste

Na generalidade dos atributos a sua previsão tem como base o seu histórico de variação, isto é, as suas previsões não vão ser influenciadas por outras variáveis a não ser pela mesma. Desta forma, todos os atributos dos 3 grupos definidos foram alvo de previsões, para que os dados se encontram todos normalizados, tendo como objetivo a definição de uma carteira de ativos que vá de encontro com os objetivos definidos.

Contudo, no caso da variável S&P500 procedeu-se ao desenvolvimento de três metodologias de previsão. Uma das metodologias, conforme referido anteriormente, é apenas com base no histórico de variação do mesmo, a segunda metodologia já integra todos os atributos do grupo S&P500 para a previsão deste atributo, exceto o histórico de variação do mesmo e a terceira metodologia é similar à segunda, contudo contempla no conjunto de dados a variação do S&P500 no passado o que faz com que seja uma mistura da primeira e segunda metodologia. Deste modo, o objetivo pretendido é verificar qual a metodologia mais fiável a utilizar no projeto para assim alcançar resultados mais fiáveis. Assim, de uma forma mais detalhada para a aplicação da segunda e da terceira metodologia de previsão do S&P500 vão ser utilizados adicionalmente as seguintes variáveis:

- EUR_USD – Variação de mercado da Relação entre o dólar e o euro;
- GDP USA – Variação de mercado do Produto Interno Bruto dos Estados Unidos da América;
- Taxas de Juro USA - Variação de mercado das Taxas de Juro dos Estados Unidos da América;
- Gold - Variação de mercado do Ouro.

Depois foi necessário proceder à divisão do *dataset* em dados de treino e em dados de teste, deste modo procedeu-se, com o auxílio do nó “*Partitioning*”, à divisão em dois conjuntos com 70% e 30% dos dados, respetivamente. No apêndice D é apresentado a parametrização do nó “*Partitioning*” utilizada.

3.4.3. Construção de Modelos

Como referido anteriormente os conjuntos de dados foram subdivididos em 2 conjuntos de treino e de teste, assim é expectável que os modelos utilizados utilizem o conjunto de treino para aprender o comportamento dos dados, de forma que depois realizam testes com o conjunto de testes. Depois, tendo em vista a previsão do mês seguinte tendo em consideração os 4 meses anteriores, procedeu-se à transformação do conjunto de dados para que o mesmo listasse os 4 meses anteriores em cada linha do *dataset*.

Para isso utilizou-se previamente o *node* “*Window Slider*”, tendo em vista a obtenção dos dados de T-3, T-2, T-1, T e T+1, sendo que T+1 representa o “*Target*” das previsões. De realçar que quando se tratou de previsões com diferentes atributos foi necessário utilizar o *node* “*Joiner*”, para integrar todos os dados no mesmo conjunto de dados. Posteriormente, consoante a previsão efetuada e o *target* definido foram avaliados e comparados os diferentes resultados dos modelos através do *node* “*Numeric Scorer*”.

3.4.4. Avaliação de Modelos

Para a avaliação dos modelos foi tida em conta a variação dos valores entre o *target* e o previsto pelos modelos, tendo como objetivo identificar qual o modelo que se adequa mais a cada um dos atributos previstos. Para isso foram utilizadas as métricas de coeficiente de determinação R^2 e os valores de erro quadrático médio (EQM) obtidos do *node* “*Numeric Scorer*”.

Tabela 11 - Avaliação Modelos Grupo ETF's

| Modelo de previsão | R ² | EQM | R ² | EQM | R ² | EQM |
|-------------------------------|----------------|-----------------|----------------|---------------|----------------|----------------|
| | ETF_DIA | | ETF_EEM | | ETF_EWZ | |
| <i>Linear Regression</i> | 0,626 | 1793,504 | 0,635 | 38,527 | 0,746 | 23,669 |
| <i>Polynomial Regression</i> | 0,431 | 2727,949 | 0,643 | 37,745 | 0,715 | 26,553 |
| <i>Simple Regression Tree</i> | -0,566 | 7510,243 | 0,366 | 66,965 | 0,101 | 83,891 |
| <i>Random Forest</i> | -0,572 | 7540,651 | 0,484 | 54,532 | 0,387 | 57,155 |
| <i>Gradient Boosted Trees</i> | -0,527 | 7325,602 | 0,527 | 50,012 | 0,414 | 54,705 |
| <i>Tree Ensemble</i> | -0,573 | 7546,548 | 0,488 | 54,156 | 0,362 | 59,549 |
| | ETF_ARKG | | ETF_FCG | | ETF_ASHR | |
| <i>Linear Regression</i> | 0,848 | 94,095 | -0,194 | 70,416 | 0,542 | 29,055 |
| <i>Polynomial Regression</i> | 0,686 | 215,482 | -3,786 | 282,178 | -1,102 | 133,271 |
| <i>Simple Regression Tree</i> | -0,309 | 808,477 | -30,939 | 1883,055 | -1,303 | 146,061 |
| <i>Random Forest</i> | -0,284 | 792,739 | -28,663 | 1748,877 | -0,832 | 116,152 |
| <i>Gradient Boosted Trees</i> | -0,429 | 882,37 | -39,937 | 2413,566 | -0,307 | 82,885 |
| <i>Tree Ensemble</i> | -0,284 | 792,59 | -30,376 | 1849,862 | -1,051 | 130,074 |
| | ETF_BBRE | | ETF_COPX | | ETF_EFG | |
| <i>Linear Regression</i> | -0,475 | 432,241 | 0,662 | 23,692 | 0,57 | 172,196 |
| <i>Polynomial Regression</i> | -0,475 | 432,241 | 0,522 | 33,466 | 0,502 | 199,403 |
| <i>Simple Regression Tree</i> | 0 | 293,02 | 0,083 | 64,169 | 0,384 | 246,623 |
| <i>Random Forest</i> | 0 | 293,02 | 0,088 | 63,847 | 0,279 | 288,505 |
| <i>Gradient Boosted Trees</i> | 0 | 293,02 | -0,071 | 75,001 | 0,346 | 261,947 |
| <i>Tree Ensemble</i> | 0 | 293,02 | 0,096 | 63,317 | 0,285 | 286,353 |

Na tabela 11 apresenta o valor das métricas utilizadas para cada ativo em análise e para cada modelo de previsão utilizado. Assim, de acordo com os dados da Tabela 11, podemos verificar que na grande maioria dos atributos, o melhor modelo a aplicar consiste no *Linear Regression*. Contudo no caso do “ETF_EEM” o modelo que apresenta um coeficiente de determinação (R^2) mais próximo de 1 e um EQM mais próximo de 0, consiste no *Polynomial Regression*. Deste modo, é possível constatar que para os grupos dos ETF's os modelos de previsão que estão associados a melhores resultados são os anteriormente referidos. De realçar o caso do “ETF_BBRE” que apresenta valores muito similares a nível do nó “*Numeric Scorer*” o que pode estar associado a que o conjunto de treino não era suficiente variado para encontrar padrões dentro do conjunto de dados, o que levou a que os modelos com melhor avaliação fossem o *Simple Regression Tree*, o *Random Forest*, o *Gradient Boosted Trees* e o *Tree Ensemble*. Além disso, também de referir a situação do “ETF_FCG” que apresenta um coeficiente de correlação negativo o que ajuda a concluir que nenhum dos modelos é adequado para o presente atributo.

De uma maneira geral os atributos apresentam uma correlação moderada, apenas para as variáveis “ETF_ARKG” e “ETF_EWZ” é que foi possível constatar uma correlação mais elevada. Posto isto, na generalidade das variáveis as previsões obtidas poderão desviar-se dos valores reais.

Por último, realçar que visto que a qualidade dos modelos para os atributos “ETF_FCG” e “ETF_BBRE” era bastante reduzida decidiu-se retirar os mesmos das previsões, de modo que estes não tenham qualquer impacto no sucesso do estudo efetuado nas fases de determinação da constituição de uma carteira ótima.

Tabela 12 – Avaliação Modelos Grupo Commodities

| Modelo de previsão | R ² | EQM | R ² | EQM | R ² | EQM |
|-------------------------------|---------------------|-----------------|----------------|-------------------|----------------|----------------|
| | Crude Oil Petroleum | | Aluminum | | Beef | |
| <i>Linear Regression</i> | 0,938 | 153,118 | 0,879 | 33407,336 | 0,86 | 299,207 |
| <i>Polynomial Regression</i> | 0,079 | 2263,214 | 0,84 | 44192,343 | 0,767 | 497,948 |
| <i>Simple Regression Tree</i> | -0,686 | 4142,467 | 0,477 | 144155,742 | -0,101 | 2354,694 |
| <i>Random Forest</i> | -0,891 | 4644,995 | 0,766 | 64500,533 | -0,378 | 2947,18 |
| <i>Gradient Boosted Trees</i> | -0,749 | 4297,107 | 0,762 | 65471,877 | -0,366 | 2920,943 |
| <i>Tree Ensemble</i> | -0,887 | 4634,13 | 0,757 | 66866,726 | -0,349 | 2884,639 |
| | Coal | | Copper | | Cotton | |
| <i>Linear Regression</i> | 0,912 | 94,468 | 0,87 | 418002,105 | 0,919 | 95,007 |
| <i>Polynomial Regression</i> | 0,914 | 92,019 | 0,548 | 1454005,63 | 0,801 | 231,727 |
| <i>Simple Regression Tree</i> | -0,755 | 1884,347 | -2,898 | 12550839,49 | 0,54 | 536,492 |
| <i>Random Forest</i> | -0,572 | 1688,614 | -3,323 | 13919361,49 | 0,515 | 565,214 |
| <i>Gradient Boosted Trees</i> | -0,524 | 1636,444 | -4,094 | 16402534,42 | 0,517 | 563,255 |
| <i>Tree Ensemble</i> | -0,55 | 1664,556 | -3,537 | 14609034,46 | 0,517 | 563,861 |
| | Rice | | Tea | | Uranium | |
| <i>Linear Regression</i> | 0,885 | 2790,177 | 0,794 | 1006,594 | 0,906 | 41,923 |
| <i>Polynomial Regression</i> | 0,736 | 6386,547 | 0,773 | 1106,924 | 0,786 | 95,025 |
| <i>Simple Regression Tree</i> | 0,559 | 10666,089 | 0,63 | 1810,099 | -0,669 | 741,766 |
| <i>Random Forest</i> | 0,51 | 11850,176 | 0,648 | 1719,16 | -0,691 | 751,892 |
| <i>Gradient Boosted Trees</i> | 0,476 | 12656,988 | 0,661 | 1657,043 | -0,675 | 744,768 |
| <i>Tree Ensemble</i> | 0,492 | 12271,985 | 0,632 | 1799,655 | -0,697 | 754,543 |
| | Wheat | | | | | |
| <i>Linear Regression</i> | 0,87 | 572,99 | | | | |
| <i>Polynomial Regression</i> | 0,846 | 677,99 | | | | |
| <i>Simple Regression Tree</i> | 0,328 | 2951,22 | | | | |
| <i>Random Forest</i> | 0,36 | 2813,666 | | | | |
| <i>Gradient Boosted Trees</i> | 0,216 | 3446,232 | | | | |
| <i>Tree Ensemble</i> | 0,305 | 3053,975 | | | | |

Com base na tabela 12 é possível destacar que na grande maioria dos atributos o modelo de previsão que possui uma melhor avaliação trata-se do “*Linear Regression*”, contudo no caso da variável do carvão o modelo “*Polynomial Regression*” é o que se define como mais adequado para a previsão da mesma dado apresentar um maior valor de coeficiente de correlação e um menor erro quadrático médio.

Ao contrário dos atributos do grupo dos *ETF's*, as variáveis em estudo no grupo das *commodities* já apresentam modelos de previsão com um elevado coeficiente de correlação, o que naturalmente está associado a uma melhor previsão e conseqüentemente menor enviesamento dos valores reais.

Tabela 13 – Avaliação Modelos ETFs

| Modelo de previsão | R ² | EQM | R ² | EQM | R ² | EQM |
|-------------------------------|----------------|------------------|----------------|------------|----------------|-----------|
| | S&P500 (1) | | S&P500 (2) | | S&P500 (3) | |
| <i>Linear Regression</i> | 0,954 | 18559,552 | -21,759 | 6334485,46 | -2,082 | 857774,27 |
| <i>Polynomial Regression</i> | 0,776 | 90067,285 | -1,884 | 802772,207 | -0,589 | 442331,47 |
| <i>Simple Regression Tree</i> | -1,813 | 1129818,77 | -0,946 | 541520,601 | -1,18 | 606655,76 |
| <i>Random Forest</i> | -1,781 | 1116809,14 | -1,578 | 717640,686 | -1,187 | 608645,56 |
| <i>Gradient Boosted Trees</i> | -1,774 | 1114198,22 | -1,175 | 605409,476 | -0,938 | 539268,66 |
| <i>Tree Ensemble</i> | -1,78 | 1116655,84 | -1,604 | 724842,215 | -1,225 | 619382,38 |

Conforme referido anteriormente para a previsão do S&P500 foram utilizadas 3 diferentes conjuntos de dados, no caso do número 1 consiste em apenas dados históricos S&P500, depois no número 2 fora utilizadas os atributos do grupo S&P500 com a exceção da variação histórica do S&P500 e, por último, no número 3 o conjunto de dados integra os dados do conjunto 1 e 2.

Posto isto, tendo por base os valores definidos na tabela 13 é possível constatar que o conjunto de dados número 1 com o modelo de previsão “*Linear Regression*” se caracteriza pela previsão com melhores resultados a nível do coeficiente de determinação e do erro quadrático médio.

De realçar que neste projeto o principal objetivo não se caracterizava pela resolução de um problema de previsão, o que levou a que não fosse desenvolvida uma análise dos parâmetros ótimos para cada modelo.

3.5. Otimização Carteira de Ativos

Para a determinação da carteira de ativos que vai de encontro com os objetivos pré-estabelecidos foram utilizada duas metodologias de forma que seja possível concluir após o estudo efetuado, qual a que se adequa melhor ao pretendido. Uma das metodologias consiste no AHP que é utilizado para auxiliar na tomada de decisão por via matemática e lógica, na mesma é possível definir diferentes atributos que sejam significantes para a tomada de decisão e para definir um valor para os mesmos consoante as diferentes alternativas definidas. Nesta metodologia é importante que o utilizador indique qual o peso na tomada de decisão de cada atributo. Deste modo, o objetivo com esta metodologia é encontrar a percentagem recomendada para cada ativo em análise para assim alcançar a carteira de ativos ótima, tendo em conta os atributos e o peso dos mesmos. Além desta metodologia, vai ser também desenvolvido um estudo de uma carteira de ativos em que seja definida a mesma percentagem para todos, isto é, tendo 10 ativos disponíveis para integrarem uma carteira de ativos, será definida uma percentagem de 10% para cada um. Assim para este estudo para o método AHP vão ser utilizados dois atributos, o risco e o lucro sendo que o objetivo é minimizar o risco e maximizar o lucro. Por fim, pretende-se efetuar uma análise das metodologias, tendo em vista a identificação da metodologia que contribui para um maior sucesso a nível de investimentos.

Posto isto, procedeu-se ao desenvolvimento de uma ferramenta em Excel capaz de calcular a percentagem de variação de cada ativo, de forma a ir de encontro com o objetivo de maximizar os lucros e minimizar o risco. Posteriormente foi necessário definir uma data a título de exemplo, de forma que seja possível efetuar uma comparação entre o método AHP e o modelo da divisão igualitária dos ativos. Para isso, foram analisados os conjuntos de dados e visto que o grupo das *Commodities* apenas possuía dados até ao início do ano de 2016, foi escolhido o mês de novembro de 2015 para efetuarmos previsões a nível de todo os ativos em análise. De salientar, como referido anteriormente, que não foram consideradas as previsões do “ETF_FCG” e do “ETF_BBRE” para a constituição de uma carteira de ativos.

Tendo em vista a obtenção da percentagem do lucro previsto foi utilizado o nó “*Math Formula*” para calcular a percentagem de lucro prevista tendo em conta a previsão do mês de novembro e o valor do ativo no mês de outubro. A configuração do nó “*Math Formula*” encontra-se definida no apêndice E. Apenas foi aplicado o nó em causa, no modelo com melhor avaliação para cada atributo analisado, conforme apresentado no apêndice F. Além do lucro, também foi necessário proceder ao cálculo do risco associado aos ativos, para isso efetuou-se a divisão do desvio-padrão pela média de cada ativo, de modo a normalizar os diferentes riscos dos atributos. O que resultou na utilização dos nós “*Date&Time-based Row Filter*” e “*Statistics*”, para filtrar os dados entre janeiro de 2000 e outubro de 2015 e calcular o desvio-padrão e a média do conjunto de dados, respetivamente, conforme apresentado no apêndice G. Na tabela 14 seguinte encontram-se definidas as previsões de lucro para o mês de novembro de 2015 e assim como o risco associado a cada ativo.

Tabela 14a – Previsão de Lucro e Risco dos Ativos

| Ativo | % Previsão Lucro mês 11/2015 | % Risco |
|---------------------|------------------------------|---------|
| ETF_DIA | 0,835% | 21,628% |
| ETF_EEM | 1,421% | 43,435% |
| ETF_EWZ | -0,887% | 57,928% |
| ETF_ARKG | -2,238% | 2,693% |
| ETF_ASHR | 9,162% | 14,427% |
| ETF_COPX | 5,170% | 19,526% |
| ETF_EFG | 0,076% | 14,800% |
| Crude Oil Petroleum | -0,494% | 48,337% |
| Aluminum | -0,672% | 23,580% |
| Beef | -3,859% | 32,136% |
| Coal | -2,905% | 51,125% |
| Copper | -1,101% | 50,517% |
| Cotton | 0,367% | 40,721% |
| Rice | -0,529% | 44,807% |
| Tea | -3,078% | 23,864% |

Tabela 14b – Previsão de Lucro e Risco dos Ativos

| Ativo | % Previsão Lucro mês 11/2015 | % Risco |
|---------|------------------------------|---------|
| Uranium | -0,428% | 65,843% |
| Wheat | -0,373% | 36,743% |
| S&P500 | 1,345% | 23,785% |

Os resultados apresentados na tabela 14 foram utilizados para que o método AHP conseguisse definir uma matriz normalizada de valores para cada ativo, para assim simplificar o cálculo das percentagens dos ativos da carteira de investimentos e de modo a ir de encontro com os objetivos definidos. Assim que o modelo foi construído com esta informação foram desenvolvidos os cálculos matriciais associados ao método AHP para o lucro e o risco que se caracterizam como os únicos atributos, estes que se encontram definidos nos apêndices H e I, respetivamente. De realçar que, visto que, o objetivo consiste em maximizar o lucro e minimizar o risco, a nível das percentagens do risco efetuou-se o cálculo inverso dos valores alcançados. Deste modo os ativos com um maior risco associado possuíam um menor coeficiente. Depois, após a normalização dos ativos, que se definem como alternativas, tendo em conta os atributos de risco e lucro foi necessário incorporar no modelo a opção de definir percentagens de significância diferentes entre os dois atributos utilizados. Contudo para este estudo foi considerada uma percentagem de 30% para o lucro e 70% para o risco, o que significa que o modelo atribui uma maior relevância à minimização do risco relativamente à maximização do lucro. Assim após a execução do método encontram-se definidos na figura 23 os resultados alcançados.

| Percentagem de Significância | | Solução | | Previsão | |
|------------------------------|-------------|---------------------|-------------|----------|--------|
| Lucro | Risco | Ativo | Percentagem | Lucro | Risco |
| 30% | 70% | ETF DIA | 17% | 25,76% | 20,44% |
| | | ETF EEM | 25% | | |
| | | ETF EWZ | -13% | | |
| | | ETF_ARKG | -10% | | |
| | | ETF_ASHR | 157% | | |
| | | ETF_COPX | 89% | | |
| | | ETF_EFG | 6% | | |
| | | Crude Oil Petroleum | -7% | | |
| | | Aluminum | -8% | | |
| | | Beef | -62% | | |
| | | Coal | -47% | | |
| | | Copper | -17% | | |
| | | Cotton | 8% | | |
| | | Rice | -7% | | |
| | | Tea | -48% | | |
| | | Uranium | -6% | | |
| | | Wheat | -4% | | |
| | | S&P500 | 25% | | |
| Total | 100% | Total | 100% | | |

Figura 23 – Resultados Otimização Carteira de Ativos através do Método AHP

Com base na análise da figura 23 é possível constatar que com estas alterações à carteira de ativos estava previsto o portefólio aumentar 25,76%, com um aumento do risco associado de 20,44%. Contudo, caso se considere 5,56% para cada ativo, isto é, utilizando o método da divisão igualitária dos ativos seria previsto alcançar um lucro de 0,10% associado a uma diminuição do risco no valor de 34,22%, conforme é possível verificar na figura 24.

| Solução | | Previsão | |
|---------------------|-------------|----------|---------|
| Ativo | Percentagem | Lucro | Risco |
| ETF_DIA | 5,56% | 0,10% | -34,22% |
| ETF_EEM | 5,56% | | |
| ETF_EWZ | 5,56% | | |
| ETF_ARKG | 5,56% | | |
| ETF_ASHR | 5,56% | | |
| ETF_COPX | 5,56% | | |
| ETF_EFG | 5,56% | | |
| Crude Oil Petroleum | 5,56% | | |
| Aluminum | 5,56% | | |
| Beef | 5,56% | | |
| Coal | 5,56% | | |
| Copper | 5,56% | | |
| Cotton | 5,56% | | |
| Rice | 5,56% | | |
| Tea | 5,56% | | |
| Uranium | 5,56% | | |
| Wheat | 5,56% | | |
| S&P500 | 5,56% | | |
| Total | 100% | | |

Figura 24 – Resultados Otimização Carteira de Ativos através do Método de divisão igualitária

Posto isto, tendo em vista a comparação entre os diferentes métodos utilizados procedeu-se à verificação da percentagem de lucro real do mês de outubro de 2015 para o mês de novembro de 2015. Assim, após a recolha dos dados foram obtidos os seguintes resultados relativamente à percentagem de lucro real, caracterizados na tabela 15.

Tabela 15a - Previsão e Realidade de Lucro do mês em estudo

| Ativo | % Previsão Lucro mês 11/2015 | % Real Lucro mês 11/2015 |
|---------------------|------------------------------|--------------------------|
| ETF_DIA | 0,835% | 9,619% |
| ETF_EEM | 1,421% | 7,888% |
| ETF_EWZ | -0,887% | 6,837% |
| ETF_ARKG | -2,238% | 10,429% |
| ETF_ASHR | 9,162% | 7,855% |
| ETF_COPX | 5,170% | 15,862% |
| ETF_EFG | 0,076% | 7,953% |
| Crude Oil Petroleum | -0,494% | -8,186% |
| Aluminum | -0,672% | -3,204% |
| Beef | -3,859% | -4,050% |
| Coal | -2,905% | 0,493% |
| Copper | -1,101% | -7,979% |
| Cotton | 0,367% | 0,281% |
| Rice | -0,529% | -0,575% |
| Tea | -3,078% | -9,715% |
| Uranium | -0,428% | -2,762% |
| Wheat | -0,373% | -4,624% |

Tabela 15b - Previsão e Realidade de Lucro do mês em estudo

| Ativo | % Previsão Lucro mês 11/2015 | % Real Lucro mês 11/2015 |
|--------|------------------------------|--------------------------|
| S&P500 | 1,345% | 2,756% |

Tendo em vista a comparação entre as duas metodologias com valores reais, procedeu-se ao cálculo do lucro real através das percentagens dos ativos de cada carteira, para isso foi necessário realizar a multiplicação da percentagem definida na carteira e a percentagem de lucro real para cada ativo. Assim, tendo em conta os valores reais apresentados a tabela 15 e as percentagens das carteiras definidas na figura 23 e 24, foi possível constatar que utilizando o método AHP podemos obter um lucro de 38,87% e o método de divisão igualitária dos ativos apenas 1,60%. De notar que apesar da carteira de ativos definida na figura 24 se traduzir numa taxa de rentabilidade menor, esta também possuía um menor risco para o investidor, assim sendo, caso o mesmo selecione um variado e diversificado conjunto de ativos e que preveja que estes vão rentabilizar a curto, médio ou longo prazo pode utilizar este método de forma que seja menos afetado pelas normais oscilações do mercado.

Em contrapartida, caso o investidor pretenda ter poder sobre o grau de importância entre o lucro e o risco e deste modo construir uma carteira de ativos que maximize a rentabilidade e diminua o risco associado a este investimento deve optar pela utilização do método AHP. Em suma, neste exemplo prático ambos os modelos conseguiram originar uma carteira de ativos com uma variação dos ativos positiva, contudo o método que não garante com tanta certeza a rentabilidade do portefólio trata-se do modelo de divisão de pesos igualitária dos ativos. Assim sendo, constata-se que o método AHP caracteriza-se pelo modelo mais completo para a definição de uma carteira de ativos que tem em conta todos os objetivos pré-estabelecidos para este estudo.

4. CONCLUSÃO

Neste capítulo serão apresentadas as principais conclusões alcançadas através deste estudo, assim como os objetivos que foram alcançados e qual a utilidade deste projeto para o futuro. Além disso, serão também retratadas as principais limitações do projeto e os próximos desafios que poderão levar ao desenvolvimento de uma ferramenta mais completa.

4.1. Conclusões finais

Um dos principais objetivos deste projeto consistia em simplificar o investimento das pessoas que não possuem grande conhecimento a nível de literacia financeira. Deste modo, seria possível que as mesmas realizassem investimentos que vão de encontro com os seus objetivos. Além de auxiliar a sociedade em geral acredito que esta ferramenta poderia ser bastante útil para fundos de investimento e bancos que possuem carteiras de ativos e que têm como objetivo a maximização do lucro e a minimização do risco associado.

Depois, em termos pessoais, este projeto potenciou o meu conhecimento financeiro com excelentes bases teóricas que contribuíram ativamente para o sucesso do mesmo. Posto isto, acredito que esta ferramenta se aproxima bastante do futuro a curto/médio prazo dos investimentos e que esta pode vingar no mercado com o devido acompanhamento.

Relativamente ao *Machine Learning* o estudo desenvolvido não assentava sobre a otimização dos modelos de previsão, porém acredito que o objetivo da previsão dos ativos foi alcançado. Esta pode ser cada vez mais reconhecida como uma hipótese viável de prever ativos financeiros sem ser necessário efetuar qualquer cálculo de previsões. A grande vantagem desta área é que permite anexar aos conjuntos de dados variáveis que apresentam uma correlação com a principal variável a ser prevista, o que potencia a otimização e exatidão dos valores previstos.

Depois, a nível das metodologias de definição da constituição da carteira de ativos foi possível concluir que o método AHP se definia como a solução mais completa para os objetivos do projeto e tendo em conta as necessidades do utilizador. A principal vantagem associada a esta metodologia consiste na sua flexibilidade em adicionar novos atributos e alternativas e que consequentemente tornam a ferramenta mais completa.

Em suma, foi possível perceber toda a complexidade associada ao mercado de ações e à constituição de uma carteira de ativos que maximize o lucro e minimize o risco associado. Além disso, é de salientar o contributo que este pode ter para sociedade em geral, tanto como utilizadores individuais da ferramenta como a nível de fundos de investimento e bancos. Posto isto, é possível concluir que o projeto foi desenvolvido com sucesso e mostrou-se proveitoso para mim pelo conhecimento ganho a nível de literacia financeira e de previsões e para a sociedade pelo contributo para mais correta e otimizada forma de investimento.

4.2. Limitações e investigação futura

Dadas as limitações temporais de execução deste projeto existem diversos pontos que carecem de uma análise e de um estudo mais aprofundado. Nesta fase o objetivo consistia em construir um

fluxo geral do funcionamento da ferramenta, de forma a constatar que a mesma se caracteriza como um objeto de interesse para desenvolvimentos futuros.

De uma maneira geral, o estudo realizado pode ainda ser bastante mais aprofundado, tanto a nível de previsões como relativamente às metodologias de definição da constituição da carteira de investimentos. Como referido anteriormente, neste projeto para os modelos de previsão não foi feito um estudo para encontrar os parâmetros ótimos para cada modelo, o que podia potenciar ainda mais os resultados obtidos das previsões, visto que o principal objetivo não era resolver um problema de previsão. Assim este poderia ser um dos pontos a melhorar num próximo desenvolvimento desta ferramenta. Além do *Machine Learning* também seria possível desenvolver uma comparação com mais metodologias de otimização da carteira de ativos e também adicionar mais atributos e alternativas à metodologia AHP, o que, por conseguinte, poderia levar a resultados mais fiáveis.

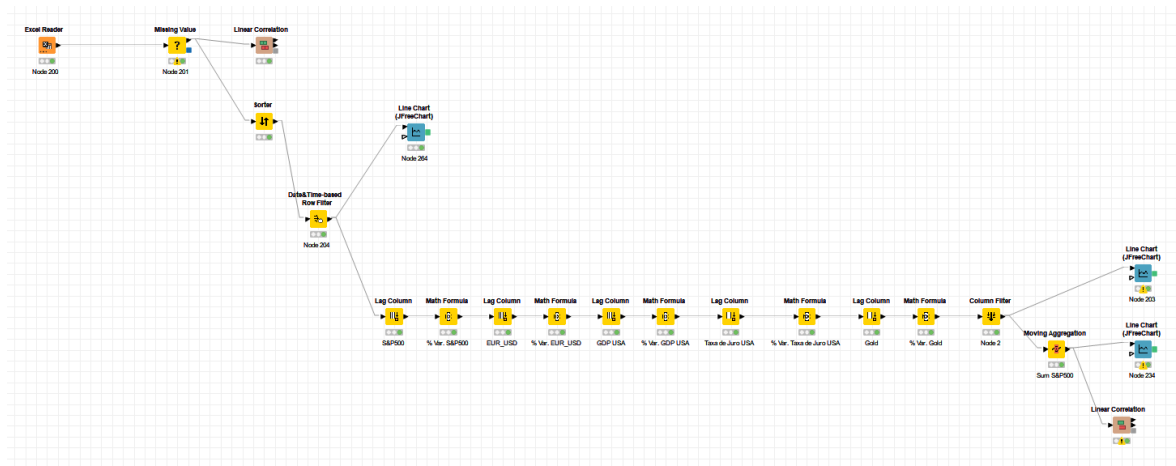
Uma das grandes limitações desta ferramenta assenta sobre a forma de utilização, que ainda se define como bastante complexa para um utilizador sem conhecimentos nestas áreas, visto que exige que sejam extraídas as previsões do Knime e depois que as mesmas sejam abastecidas no ficheiro Excel desenvolvido para definir a constituição da carteira de ativos. Posto isto, um dos próximos passos poderia passar pelo aumento da robustez e pela melhoria da experiência de utilização, tendo em vista o aumento dos utilizadores capazes de utilizar a ferramenta com sucesso.

REFERÊNCIAS BIBLIOGRÁFICAS

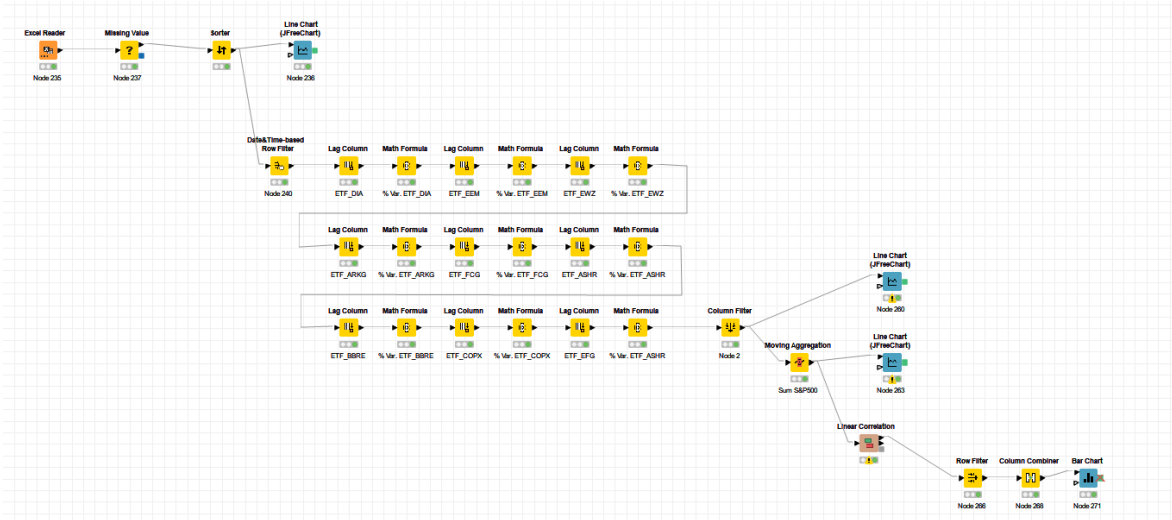
- Bakos, G. (2013). *KNIME Essentials Perform accurate data analysis using the power of KNIME*. Packt Publishing. Retrieved from www.packtpub.com
- Barsacchi, R., Carpent, A., Stöter, M., Niederlein, A., Meyenhofer, F., Brandl, H., & Bickle, M. (2013). CellProfiler and KNIME: Open Source Tools for High Content Screening. *Methods in Molecular Biology*, 986. https://doi.org/10.1007/978-1-62703-311-4_8
- Ben, F., & Ufrgs, U. C. S. (2006). Utilização do método AHP em decisões de investimento ambiental, (March), 1–8.
- Bodie, Z., Kane, A., & Marcus, A. J. (2013). *Investments*.
- Brealey, R., Myers, S., & Franklin, A. (2020). *Principles of Corporate Finance* (13th ed.). MC Graw Hill.
- Carneiro Da Rocha, B., Timoteo De Sousa Junior, R., & Timóteo De Sousa Júnior, R. (2010). Identifying Bank Frauds Using CRISP-DM and Decision Trees. *International Journal of Computer Science & Information Technology (IJCSIT)*, 2(5). <https://doi.org/10.5121/ijcsit.2010.2512>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Sherarer, C., & Wirth, R. (1999). CRISP-DM 1.0 Step-by-step data mining guide.
- Chen, J. (2021). Real Estate. Retrieved January 30, 2022, from <https://www.investopedia.com/terms/r/realstate.asp>
- Chkir, I. E., & Cosset, J. C. (2001). Diversification strategy and capital structure of multinational corporations. *Journal of Multinational Financial Management*, 11(1), 17–37. [https://doi.org/10.1016/S1042-444X\(00\)00040-2](https://doi.org/10.1016/S1042-444X(00)00040-2)
- Colpani, R. (2019). Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. *Informática Na Educação: Teoria & Prática*, 21(3), 143–157. <https://doi.org/10.22456/1982-1654.87880>
- Costa, H. G. (2002). Introdução ao método de análise hierárquica: análise multicritério no auxílio à decisão. *XXXVI Simpósio Brasileiro de Pesquisa Operacional*, 122.
- Coutinho, C. P. (2011). *Metodologia de Investigação em Ciências Sociais e Humanas: Teoria e Prática* (Edições Al).
- Couto, F. M. G. (2013). Literacia Financeira e Empreendedorismo em Portugal por Dissertação do Master in Finance.
- Degiro. (2022). O que são Ações? Retrieved January 30, 2022, from https://www.degiro.pt/conhecimento/investir-em-acoas?utm_source=google&utm_medium=cpc&utm_campaign=nonbranded&utm_content=10270159811-99755729942&gclid=CjwKCAiA55mPBhBOEiwANmzoQjDPTozWNnBm4jKl-hDq4HzGeSMO2Hok1ORp-nQH8zddcdKCDmjKxoCWmUQAvD_BwE
- Dimnson, E., Marsh, P., & Staunton, M. (2021). *Triumph of the Optimists. Angewandte Chemie International Edition*, 6(11), 951–952.
- Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2013). *Modern Portfolio Theory and Investments Analysis. 9th ed. Wiley* (Vol. 53).
- Fabozzi, F. J., & Markowitz, H. M. (2011). *The Theory and Practice of Investment Management: Asset Allocation, Valuation, Portfolio Construction, and Strategies, Second Edition. The Theory and Practice of Investment Management: Asset Allocation, Valuation, Portfolio*

- Construction, and Strategies, Second Edition*. <https://doi.org/10.1002/9781118267028>
- Fama, E. F., French, K. R., Constan-Tinides, G., Ferson, W., George, E., Harvey, C., ... Zmijewski, M. (1992). *The Cross-Section of Expected Stock Returns*. *THE JOURNAL OF FINANCE* * (Vol. XLVII). <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Feltrin, L. (2015). KNIME an Open Source Solution for Predictive Analytics in the Geosciences. <https://doi.org/10.1109/MGRS.2015.2496160>
- Ferreira, C. (2021a). CRISP-DM: A Data Mining Process.
- Ferreira, C. (2021b). CRISP-DM: Data Understanding and Data Preparation.
- Ferreira, C. (2021c). CRISP-DM: Evaluation.
- Ferreira, C. (2021d). CRISP-DM: Modeling.
- Ferreira, C. (2021e). Data Mining and Knowledge Discovery.
- Frankenfield, J. (2022). Cryptocurrency. Retrieved January 30, 2022, from <https://www.investopedia.com/terms/c/cryptocurrency.asp>
- Hayes, A. (2021a). Commodity Market . Retrieved January 30, 2022, from <https://www.investopedia.com/terms/c/commodity-market.asp>
- Hayes, A. (2021b). Dollar-Cost Averaging . Retrieved January 30, 2022, from <https://www.investopedia.com/terms/d/dollarcostaveraging.asp>
- Hayes, A. (2022). Blockchain . Retrieved January 30, 2022, from <https://www.investopedia.com/terms/b/blockchain.asp>
- Henri. (2018). *Essential of Investments*. Bodie, ZVI Kane, Alex Marcus, Alan J. MC Graw Hill.
- Investor.gov. (2021). Stocks. Retrieved January 30, 2022, from <https://www.investor.gov/introduction-investing/investing-basics/investment-products/stocks>
- Knime. (2022). KNIME Quickstart Guide. Retrieved January 26, 2022, from https://docs.knime.com/2018-12/analytics_platform_quickstart_guide/index.html#start-knime-analytics-platform
- Machado, C., & Davim, J. P. (2020). *Research Methodology in Management and Industrial Engineering*. Retrieved from <https://drive.google.com/file/d/1fZPK4BVe8cCAJfyO7g4BVsW6nMKyt8D-/view>
- Marins, C. S., Souza, D. de O., & Barros, M. da S. (2009). O Uso Do Método De Análise Hierárquica (AHP) Na Tomada De Decisões Gerenciais - Um Estudo. *Xli Sbp*, 11.
- Moro, S., Laureano, R. M. S., & Cortez, P. (2011). USING DATA MINING FOR BANK DIRECT MARKETING: AN APPLICATION OF THE CRISP-DM METHODOLOGY.
- O'hagan, S., & Kell, D. B. (2015). Software review: the KNIME workflow environment and its applications in genetic programming and machine learning. *Genetic Programming and Evolvable Machines*, 16. <https://doi.org/10.1007/s10710-015-9247-3>
- Oliveira, L. A. (2011). *Dissertação e Tese em Ciências e Tecnologia segundo Bolonha*. Lisboa: Lidel.
- Teles, D. F. (2011). *Acerca da Literacia Financeira*.
- Trevizano, W. A., & Freitas, A. L. P. (2005). Emprego do Método da Análise Hierárquica (A.H.P.) na seleção de processadores. *XXVI Encontro Nacional de Engenharia de Produção*, 2972–2979.
- Zhang, H. (2004). The Optimality of Naive Bayes. Retrieved from www.aaai.org

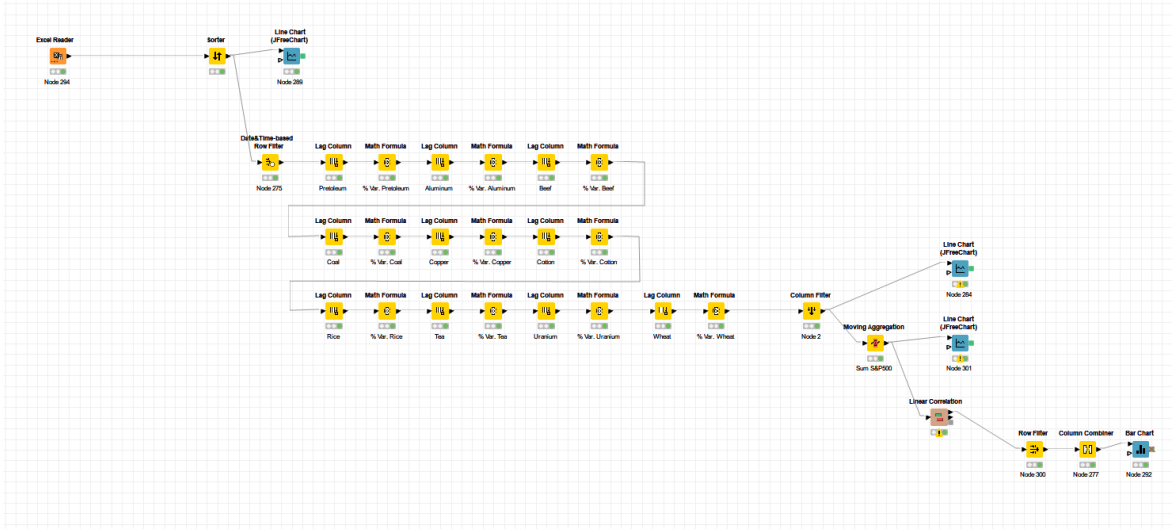
APÊNDICE A



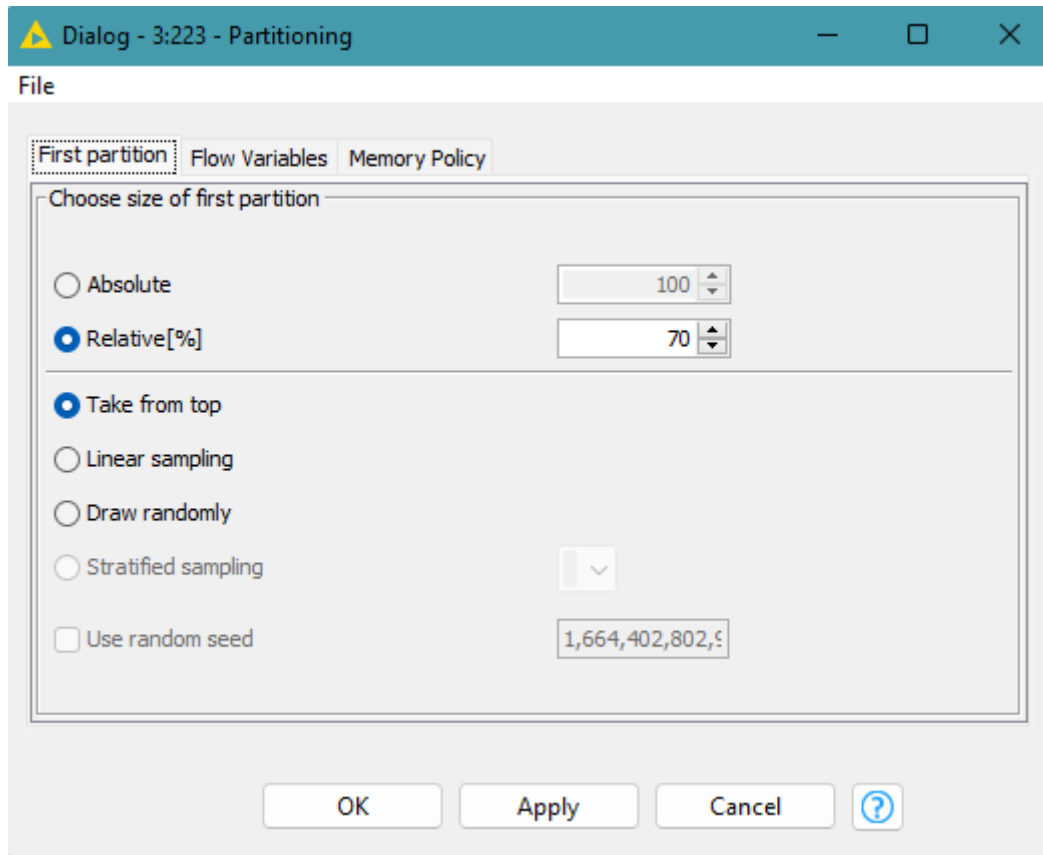
APÊNDICE B



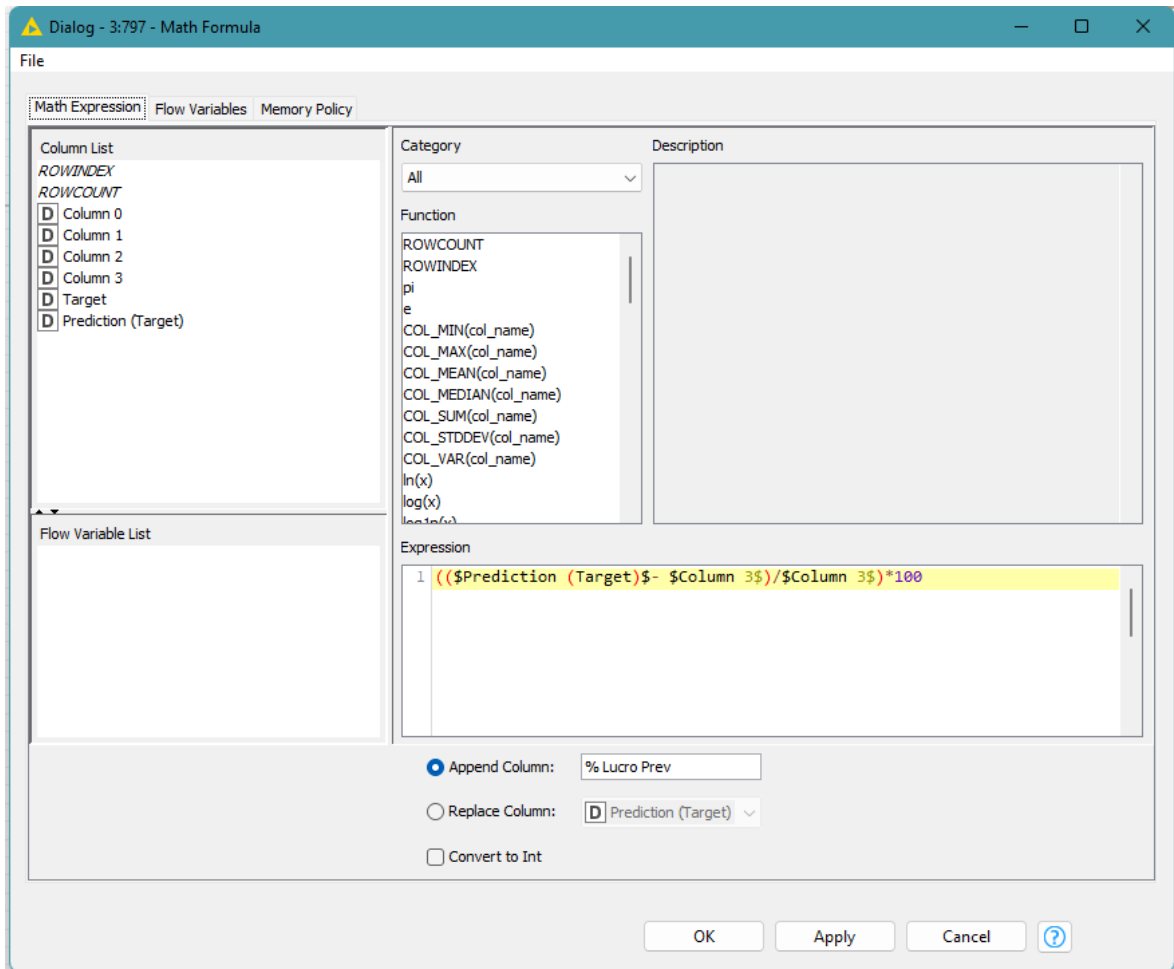
APÊNDICE C



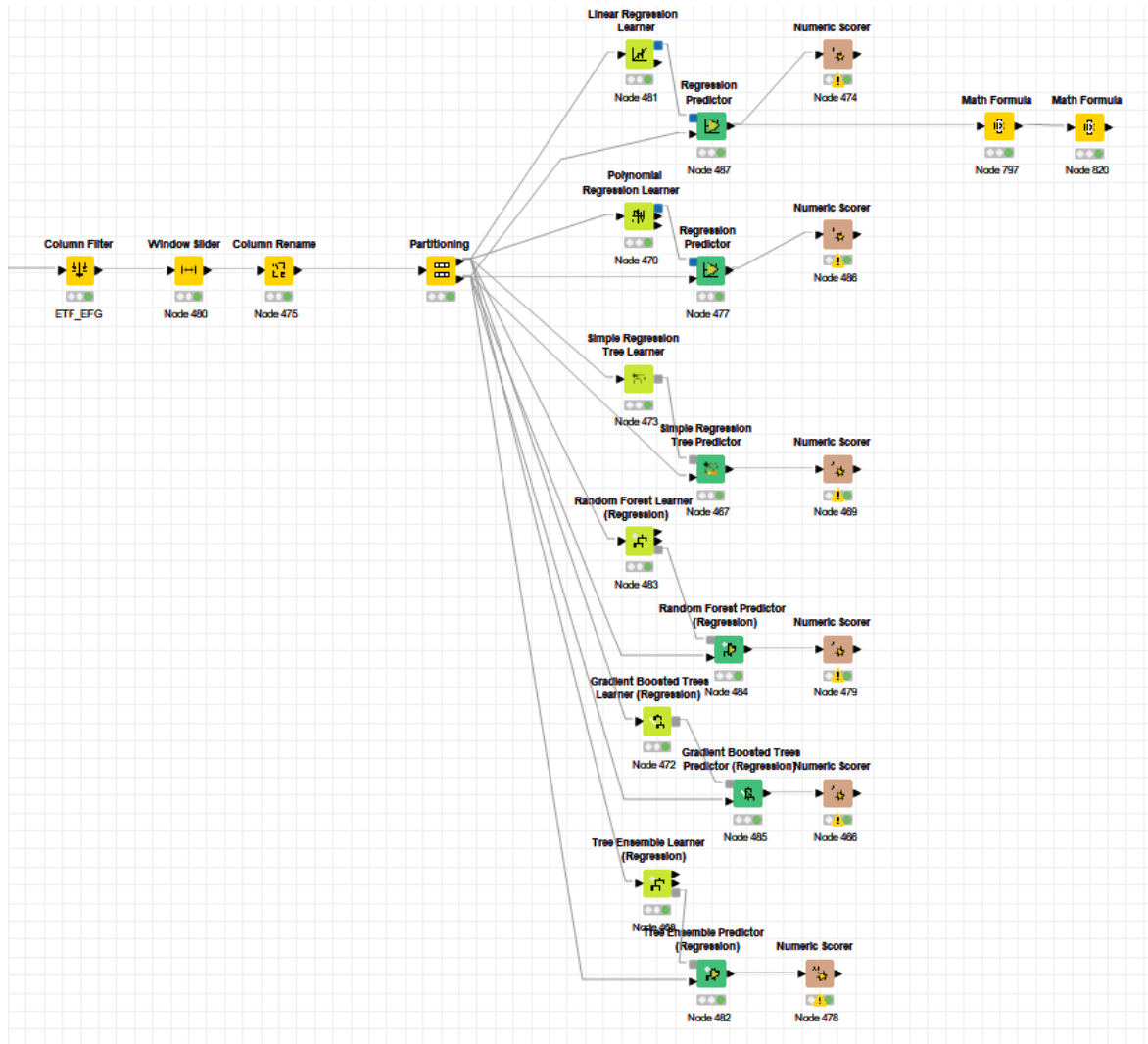
APÊNDICE D



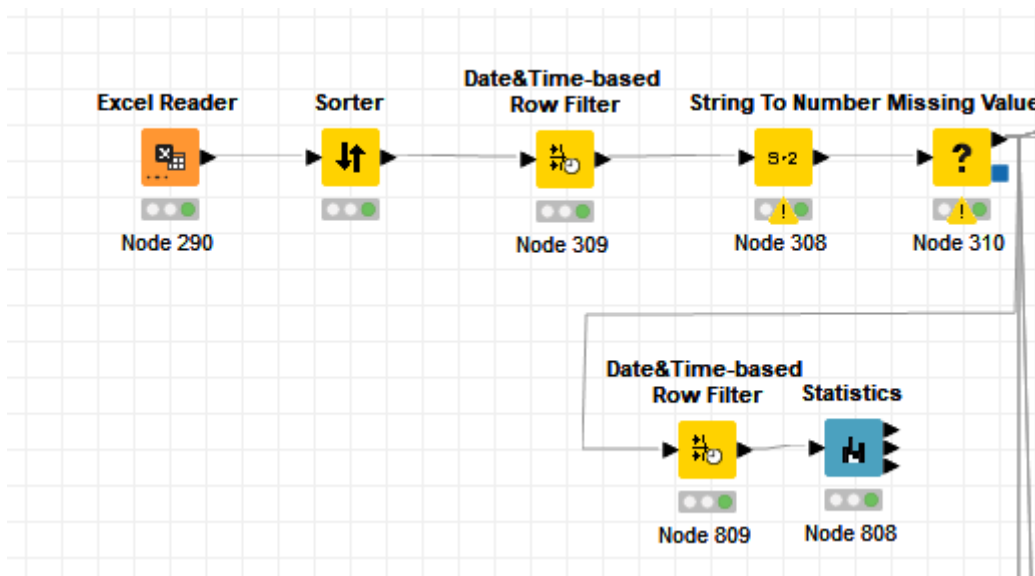
APÊNDICE E



APÊNDICE F



APÊNDICE G



APÊNDICE H

| Lucro | ETF_DIA | ETF_EEM | ETF_EWZ | ETF_ARKG | ETF_ASHR | ETF_COPX | ETF_EFG | Crude Oil Petroleum | Aluminum | Beef | Coal | Copper | Cotton | Rice | Tea | Uranium | Wheat | S&P500 | Lucro | Realidade |
|---------------------|---------|---------|---------|----------|----------|----------|---------|---------------------|----------|-------|-------|--------|--------|--------|-------|---------|--------|--------|---------|-----------|
| ETF_DIA | 1,00 | 0,59 | -0,94 | -0,37 | 0,09 | 0,16 | 10,99 | -1,69 | -1,24 | -0,22 | -0,29 | -0,76 | 2,28 | -1,58 | -0,27 | -1,95 | -2,24 | 0,62 | 0,855% | 9,519% |
| ETF_EEM | 1,70 | 1,00 | -1,60 | -0,63 | 0,16 | 0,27 | 18,70 | -2,88 | -2,11 | -0,37 | -0,49 | -1,29 | 3,87 | -2,69 | -0,46 | -3,32 | -3,81 | 1,06 | 1,421% | 7,888% |
| ETF_EWZ | -1,06 | -0,62 | 1,00 | 0,40 | -0,10 | -0,17 | -11,67 | 1,80 | 1,32 | 0,23 | 0,31 | 0,81 | -2,42 | 1,68 | 0,29 | 2,07 | 2,38 | -0,66 | -0,887% | 6,837% |
| ETF_ARKG | -2,68 | -1,57 | 2,52 | 1,00 | -0,24 | -0,43 | -29,45 | 4,53 | 3,33 | 0,58 | 0,77 | 2,03 | -6,10 | 4,23 | 0,73 | 5,23 | 6,00 | -1,66 | -2,238% | 10,429% |
| ETF_ASHR | 10,97 | 6,45 | -10,33 | -4,09 | 1,00 | 1,77 | 120,55 | -18,55 | -13,63 | -2,37 | -3,15 | -8,32 | 24,96 | -17,32 | -2,98 | -21,41 | -24,56 | 6,81 | 9,162% | 7,855% |
| ETF_COPX | 6,19 | 3,64 | -5,83 | -2,31 | 0,56 | 1,00 | 68,03 | -10,47 | -7,69 | -1,34 | -1,78 | -4,70 | 14,09 | -9,77 | -1,68 | -12,08 | -13,86 | 3,84 | 5,170% | 15,862% |
| ETF_EFG | 0,09 | 0,05 | -0,09 | -0,03 | 0,01 | 0,01 | 1,00 | -0,15 | -0,11 | -0,02 | -0,03 | -0,07 | 0,21 | -0,14 | -0,02 | -0,18 | -0,20 | 0,06 | 0,076% | 7,953% |
| Crude Oil Petroleum | -0,59 | -0,35 | 0,56 | 0,22 | -0,05 | -0,10 | -6,50 | 1,00 | 0,74 | 0,13 | 0,17 | 0,45 | -1,35 | 0,93 | 0,16 | 1,15 | 1,32 | -0,37 | -0,494% | -8,186% |
| Aluminum | -0,80 | -0,47 | 0,76 | 0,30 | -0,07 | -0,13 | -8,84 | 1,36 | 1,00 | 0,17 | 0,23 | 0,61 | -1,83 | 1,27 | 0,22 | 1,57 | 1,80 | -0,50 | -0,672% | -3,204% |
| Beef | -4,62 | -2,72 | 4,35 | 1,72 | -0,42 | -0,75 | -50,78 | 7,81 | 5,74 | 1,00 | 1,33 | 3,50 | -10,51 | 7,29 | 1,25 | 9,02 | 10,35 | -2,87 | -3,859% | -4,050% |
| Coal | -3,48 | -2,04 | 3,28 | 1,30 | -0,32 | -0,56 | -38,22 | 5,88 | 4,32 | 0,75 | 1,00 | 2,64 | -7,92 | 5,49 | 0,94 | 6,79 | 7,79 | -2,16 | -2,905% | 0,493% |
| Copper | -1,32 | -0,77 | 1,24 | 0,49 | -0,12 | -0,21 | -14,49 | 2,23 | 1,64 | 0,29 | 0,38 | 1,00 | -3,00 | 2,08 | 0,36 | 2,57 | 2,95 | -0,82 | -1,101% | -7,979% |
| Cotton | 0,44 | 0,26 | -0,41 | -0,16 | 0,04 | 0,07 | 4,83 | -0,74 | -0,55 | -0,10 | -0,13 | -0,33 | 1,00 | -0,69 | -0,12 | -0,86 | -0,98 | 0,27 | 0,367% | 0,281% |
| Rice | -0,63 | -0,37 | 0,60 | 0,24 | -0,06 | -0,10 | -6,96 | 1,07 | 0,79 | 0,14 | 0,18 | 0,48 | -1,44 | 1,00 | 0,17 | 1,24 | 1,42 | -0,39 | -0,529% | -0,575% |
| Tea | -3,69 | -2,17 | 3,47 | 1,38 | -0,34 | -0,60 | -40,50 | 6,23 | 4,58 | 0,80 | 1,06 | 2,80 | -8,39 | 5,82 | 1,00 | 7,19 | 8,25 | -2,29 | -3,078% | -9,715% |
| Uranium | -0,51 | -0,30 | 0,48 | 0,19 | -0,05 | -0,08 | -5,63 | 0,87 | 0,64 | 0,11 | 0,15 | 0,39 | -1,17 | 0,81 | 0,14 | 1,00 | 1,15 | -0,32 | -0,428% | -4,282% |
| Wheat | -0,45 | -0,26 | 0,42 | 0,17 | -0,04 | -0,07 | -4,91 | 0,76 | 0,56 | 0,10 | 0,13 | 0,34 | -1,02 | 0,71 | 0,12 | 0,87 | 1,00 | -0,28 | -0,373% | -6,624% |
| S&P500 | 1,61 | 0,95 | -1,52 | -0,60 | 0,15 | 0,26 | 17,70 | -2,72 | -2,00 | -0,35 | -0,46 | -1,22 | 3,66 | -2,54 | -0,44 | -3,14 | -3,61 | 1,00 | 1,345% | 2,756% |
| Soma | 2,17 | 1,28 | -2,04 | -0,81 | 0,20 | 0,35 | 23,84 | -3,67 | -2,70 | -0,47 | -0,62 | -1,65 | 4,94 | -3,43 | -0,59 | -4,23 | -4,86 | 1,35 | | |

| ETF_DIA | ETF_EEM | ETF_EWZ | ETF_ARKG | ETF_ASHR | ETF_COPX | ETF_EFG | Crude Oil Petroleum | Aluminum | Beef | Coal | Copper | Cotton | Rice | Tea | Uranium | Wheat | S&P500 | Total | Average |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------|-------------|
| 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 0,460816777 | 8,294 | 0,460816777 |
| 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 0,784216336 | 14,1159 | 0,784216336 |
| -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -0,48951435 | -8,1113 | -0,48951435 |
| -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -1,23509934 | -22,232 | -1,23509934 |
| 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 5,056291391 | 91,0132 | 5,056291391 |
| 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 2,853200883 | 51,3576 | 2,853200883 |
| 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,041942605 | 0,75497 | 0,041942605 |
| -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -0,37262693 | -4,8073 | -0,37262693 |
| -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -0,37086093 | -6,7555 | -0,37086093 |
| -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -2,12969095 | -38,334 | -2,12969095 |
| -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -1,60320088 | -28,858 | -1,60320088 |
| -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -0,60761589 | -10,937 | -0,60761589 |
| 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 0,202538631 | 3,6457 | 0,202538631 |
| -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -0,2919426 | -5,255 | -0,2919426 |
| -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -1,6986755 | -30,576 | -1,6986755 |
| -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -0,23620309 | -4,2517 | -0,23620309 |
| -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -0,20584989 | -3,7053 | -0,20584989 |
| 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 0,742273731 | 13,3609 | 0,742273731 |
| 2,17 | 1,28 | -2,04 | -0,81 | 0,20 | 0,35 | 23,84 | -3,67 | -2,70 | -0,47 | -0,62 | -1,65 | 4,94 | -3,43 | -0,59 | -4,23 | -4,86 | 1,35 | | |

APÊNDICE I

| Risco | ETF_DIA | ETF_EEM | ETF_EWZ | ETF_ARKG | ETF_ASHR | ETF_COPX | ETF_EFG | Crude Oil Petroleum | Aluminum | Beef | Coal | Copper | Cotton | Rice | Tea | Uranium | Wheat | S&P500 | Risco | |
|---------------------|--------------|--------------|--------------|-------------|--------------|--------------|--------------|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------|----------|
| ETF_DIA | 1,00 | 2,01 | 2,68 | 0,12 | 0,67 | 0,90 | 0,68 | 2,23 | 1,09 | 1,49 | 2,36 | 2,34 | 1,88 | 2,07 | 1,10 | 3,04 | 1,70 | 1,10 | 4,62 | 4,623548 |
| ETF_EEM | 0,50 | 1,00 | 1,33 | 0,06 | 0,33 | 0,45 | 0,34 | 1,11 | 0,54 | 0,74 | 1,18 | 1,16 | 0,94 | 1,03 | 0,55 | 1,52 | 0,85 | 0,55 | 2,30 | 2,302268 |
| ETF_EWZ | 0,37 | 0,75 | 1,00 | 0,05 | 0,25 | 0,34 | 0,26 | 0,83 | 0,41 | 0,55 | 0,88 | 0,87 | 0,70 | 0,77 | 0,41 | 1,14 | 0,63 | 0,41 | 1,73 | 1,726295 |
| ETF_ARKG | 8,03 | 16,13 | 21,51 | 1,00 | 5,36 | 7,25 | 5,49 | 17,95 | 8,75 | 11,93 | 18,98 | 18,76 | 15,12 | 16,64 | 8,86 | 24,45 | 13,64 | 8,83 | 37,13 | 37,12663 |
| ETF_ASHR | 1,50 | 3,01 | 4,02 | 0,19 | 1,00 | 1,35 | 1,03 | 3,35 | 1,63 | 2,23 | 3,54 | 3,50 | 2,82 | 3,11 | 1,65 | 4,56 | 2,55 | 1,65 | 6,93 | 6,931558 |
| ETF_COPX | 1,11 | 2,22 | 2,97 | 0,14 | 0,74 | 1,00 | 0,76 | 2,48 | 1,21 | 1,65 | 2,62 | 2,59 | 2,09 | 2,29 | 1,22 | 3,37 | 1,88 | 1,22 | 5,12 | 5,121378 |
| ETF_EFG | 1,46 | 2,93 | 3,91 | 0,18 | 0,97 | 1,32 | 1,00 | 3,27 | 1,59 | 2,17 | 3,45 | 3,41 | 2,75 | 3,03 | 1,61 | 4,45 | 2,48 | 1,61 | 6,76 | 6,756977 |
| Crude Oil Petroleum | 0,45 | 0,90 | 1,20 | 0,06 | 0,30 | 0,40 | 0,31 | 1,00 | 0,49 | 0,66 | 1,06 | 1,05 | 0,84 | 0,93 | 0,49 | 1,36 | 0,76 | 0,49 | 2,07 | 2,06881 |
| Aluminum | 0,92 | 1,84 | 2,46 | 0,11 | 0,51 | 0,63 | 0,63 | 2,05 | 1,00 | 1,36 | 2,17 | 2,14 | 1,73 | 1,90 | 1,01 | 2,79 | 1,56 | 1,01 | 4,24 | 4,240815 |
| Beef | 0,67 | 1,35 | 1,80 | 0,08 | 0,45 | 0,61 | 0,46 | 1,50 | 0,73 | 1,00 | 1,59 | 1,57 | 1,27 | 1,39 | 0,74 | 2,05 | 1,14 | 0,74 | 3,11 | 3,111753 |
| Coal | 0,42 | 0,85 | 1,13 | 0,05 | 0,28 | 0,38 | 0,29 | 0,95 | 0,46 | 0,63 | 1,00 | 0,99 | 0,80 | 0,88 | 0,47 | 1,29 | 0,72 | 0,47 | 1,96 | 1,95959 |
| Copper | 0,43 | 0,86 | 1,15 | 0,05 | 0,29 | 0,39 | 0,29 | 0,96 | 0,47 | 0,64 | 1,01 | 1,00 | 0,81 | 0,89 | 0,47 | 1,30 | 0,73 | 0,47 | 1,98 | 1,979542 |
| Cotton | 0,53 | 1,07 | 1,42 | 0,07 | 0,35 | 0,48 | 0,36 | 1,19 | 0,58 | 0,79 | 1,26 | 1,24 | 1,00 | 1,10 | 0,59 | 1,62 | 0,90 | 0,58 | 2,46 | 2,455737 |
| Rice | 0,48 | 0,97 | 1,29 | 0,06 | 0,32 | 0,44 | 0,33 | 1,08 | 0,53 | 0,72 | 1,14 | 1,13 | 0,91 | 1,00 | 0,53 | 1,47 | 0,82 | 0,53 | 2,23 | 2,23179 |
| Tea | 0,91 | 1,82 | 2,43 | 0,11 | 0,60 | 0,82 | 0,62 | 2,03 | 0,99 | 1,35 | 2,14 | 2,12 | 1,71 | 1,88 | 1,00 | 2,76 | 1,54 | 1,00 | 4,19 | 4,190492 |
| Uranium | 0,33 | 0,66 | 0,88 | 0,04 | 0,22 | 0,30 | 0,22 | 0,73 | 0,36 | 0,49 | 0,78 | 0,77 | 0,62 | 0,68 | 0,36 | 1,00 | 0,56 | 0,36 | 1,52 | 1,51877 |
| Wheat | 0,59 | 1,18 | 1,58 | 0,07 | 0,39 | 0,53 | 0,40 | 1,32 | 0,64 | 0,87 | 1,39 | 1,37 | 1,11 | 1,22 | 0,65 | 1,79 | 1,00 | 0,65 | 2,72 | 2,712597 |
| S&P500 | 0,91 | 1,83 | 2,44 | 0,11 | 0,61 | 0,82 | 0,62 | 2,03 | 0,99 | 1,35 | 2,15 | 2,12 | 1,71 | 1,88 | 1,00 | 2,77 | 1,54 | 1,00 | 4,20 | 4,204333 |
| Soma | 20,61 | 41,38 | 55,19 | 2,57 | 13,74 | 18,60 | 14,10 | 46,05 | 22,46 | 30,62 | 48,71 | 48,13 | 38,79 | 42,69 | 22,73 | 62,73 | 35,00 | 22,66 | | |

| ETF_DIA | ETF_EEM | ETF_EWZ | ETF_ARKG | ETF_ASHR | ETF_COPX | ETF_EFG | Crude Oil Petroleum | Aluminum | Beef | Coal | Copper | Cotton | Rice | Tea | Uranium | Wheat | S&P500 | Total | Average | |
|-------------|--------------|--------------|--------------|-------------|--------------|--------------|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------|----------|
| 0,048531865 | 0,048531865 | 0,048531865 | 0,048531865 | 0,048532 | 0,0485319 | 0,04853186 | 0,048531865 | 0,0485319 | 0,048532 | 0,048532 | 0,048532 | 0,0485 | 0,0485 | 0,0485 | 0,0485 | 0,04853 | 0,0485 | 0,0485 | 0,873574 | 0,048532 |
| 0,024166155 | 0,024166155 | 0,024166155 | 0,024166155 | 0,024166 | 0,0241662 | 0,02416615 | 0,024166155 | 0,0241662 | 0,024166 | 0,024166 | 0,0242 | 0,0242 | 0,0242 | 0,0242 | 0,0242 | 0,0242 | 0,0242 | 0,0242 | 0,434991 | 0,024166 |
| 0,018120358 | 0,018120358 | 0,018120358 | 0,018120358 | 0,01812 | 0,0181204 | 0,01812036 | 0,018120358 | 0,0181204 | 0,01812 | 0,01812 | 0,0181 | 0,0181 | 0,0181 | 0,0181 | 0,0181 | 0,01812 | 0,0181 | 0,0181 | 0,326166 | 0,01812 |
| 0,389706113 | 0,389706113 | 0,389706113 | 0,389706113 | 0,389706 | 0,3897061 | 0,38970611 | 0,389706113 | 0,3897061 | 0,389706 | 0,389706 | 0,3897 | 0,3897 | 0,3897 | 0,3897 | 0,3897 | 0,38971 | 0,3897 | 0,3897 | 7,01471 | 0,389706 |
| 0,072758289 | 0,072758289 | 0,072758289 | 0,072758289 | 0,072758 | 0,0727583 | 0,07275829 | 0,072758289 | 0,0727583 | 0,072758 | 0,072758 | 0,0728 | 0,0728 | 0,0728 | 0,0728 | 0,0728 | 0,0728 | 0,0728 | 0,0728 | 1,309649 | 0,072758 |
| 0,053757429 | 0,053757429 | 0,053757429 | 0,053757429 | 0,053757 | 0,0537574 | 0,05375743 | 0,053757429 | 0,0537574 | 0,053757 | 0,053757 | 0,0538 | 0,0538 | 0,0538 | 0,0538 | 0,0538 | 0,05376 | 0,0538 | 0,0538 | 0,967634 | 0,053757 |
| 0,070925772 | 0,070925772 | 0,070925772 | 0,070925772 | 0,070926 | 0,0709257 | 0,07092577 | 0,070925772 | 0,0709257 | 0,070926 | 0,070926 | 0,0709 | 0,0709 | 0,0709 | 0,0709 | 0,0709 | 0,0709 | 0,0709 | 0,0709 | 1,276664 | 0,070926 |
| 0,021715617 | 0,021715617 | 0,021715617 | 0,021715617 | 0,021716 | 0,0217156 | 0,02171562 | 0,021715617 | 0,0217156 | 0,021716 | 0,021716 | 0,0217 | 0,0217 | 0,0217 | 0,0217 | 0,0217 | 0,02172 | 0,0217 | 0,0217 | 0,390881 | 0,021716 |
| 0,044514444 | 0,044514444 | 0,044514444 | 0,044514444 | 0,044514 | 0,0445144 | 0,04451444 | 0,044514444 | 0,0445144 | 0,044514 | 0,044514 | 0,0445 | 0,0445 | 0,0445 | 0,0445 | 0,0445 | 0,04451 | 0,0445 | 0,0445 | 0,80126 | 0,044514 |
| 0,032663056 | 0,032663056 | 0,032663056 | 0,032663056 | 0,032663 | 0,0326631 | 0,03266306 | 0,032663056 | 0,0326631 | 0,032663 | 0,032663 | 0,0327 | 0,0327 | 0,0327 | 0,0327 | 0,0327 | 0,03266 | 0,0327 | 0,0327 | 0,587935 | 0,032663 |
| 0,020531381 | 0,020531381 | 0,020531381 | 0,020531381 | 0,020531 | 0,0205314 | 0,02053138 | 0,020531381 | 0,0205314 | 0,020531 | 0,020531 | 0,0205 | 0,0205 | 0,0205 | 0,0205 | 0,0205 | 0,02053 | 0,0205 | 0,0205 | 0,369565 | 0,020531 |
| 0,020778602 | 0,020778602 | 0,020778602 | 0,020778602 | 0,020779 | 0,0207786 | 0,0207786 | 0,020778602 | 0,0207786 | 0,020779 | 0,020779 | 0,0208 | 0,0208 | 0,0208 | 0,0208 | 0,0208 | 0,02078 | 0,0208 | 0,0208 | 0,374015 | 0,020779 |
| 0,025777071 | 0,025777071 | 0,025777071 | 0,025777071 | 0,025777 | 0,0257771 | 0,02577707 | 0,025777071 | 0,0257771 | 0,025777 | 0,025777 | 0,0258 | 0,0258 | 0,0258 | 0,0258 | 0,0258 | 0,02578 | 0,0258 | 0,0258 | 0,463987 | 0,025777 |
| 0,023426366 | 0,023426366 | 0,023426366 | 0,023426366 | 0,023426 | 0,0234264 | 0,02342637 | 0,023426366 | 0,0234264 | 0,023426 | 0,023426 | 0,0234 | 0,0234 | 0,0234 | 0,0234 | 0,0234 | 0,02343 | 0,0234 | 0,0234 | 0,421675 | 0,023426 |
| 0,043986226 | 0,043986226 | 0,043986226 | 0,043986226 | 0,043986 | 0,0439862 | 0,04398623 | 0,043986226 | 0,0439862 | 0,043986 | 0,043986 | 0,044 | 0,044 | 0,044 | 0,044 | 0,04399 | 0,044 | 0,044 | 0,044 | 0,791752 | 0,043986 |
| 0,015942036 | 0,015942036 | 0,015942036 | 0,015942036 | 0,015942 | 0,015942 | 0,01594204 | 0,015942036 | 0,015942 | 0,015942 | 0,015942 | 0,0159 | 0,0159 | 0,0159 | 0,0159 | 0,0159 | 0,01594 | 0,0159 | 0,0159 | 0,288957 | 0,015942 |
| 0,028567715 | 0,028567715 | 0,028567715 | 0,028567715 | 0,028568 | 0,0285677 | 0,02856772 | 0,028567715 | 0,0285677 | 0,028568 | 0,028568 | 0,0286 | 0,0286 | 0,0286 | 0,0286 | 0,02857 | 0,0286 | 0,0286 | 0,0286 | 0,514219 | 0,028568 |
| 0,044131507 | 0,044131507 | 0,044131507 | 0,044131507 | 0,044132 | 0,0441315 | 0,04413151 | 0,044131507 | 0,0441315 | 0,044132 | 0,044132 | 0,0441 | 0,0441 | 0,0441 | 0,0441 | 0,0441 | 0,04413 | 0,0441 | 0,0441 | 0,794367 | 0,044132 |
| Soma | 20,61 | 41,38 | 55,19 | 2,57 | 13,74 | 18,60 | 14,10 | 46,05 | 22,46 | 30,62 | 48,71 | 48,13 | 38,79 | 42,69 | 22,73 | 62,73 | 35,00 | 22,66 | | |