

Perfil Invest, utilização de técnicas de agrupamento para definição de ações para estudo de acordo com perfil de investidor.

Guilherme Salgado Alves
Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo
São José dos Campos, Brasil
guilherme.salgado@unifesp.br

Leonardo Teixeira Cândido
Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo
São José dos Campos, Brasil
leonardo.teixeira@unifesp.br

Resumo— O mercado financeiro consiste na alocação de ativos entre as diversas opções de produtos oferecidos por bancos e corretoras. Considerando a imensa diversidade de opções, a melhor e talvez única possibilidade de ser bem sucedido no mercado é estar corretamente alinhado com seus próprios objetivos. Com base nesta demanda que existem os perfis de investidores do mercado financeiro. O objetivo deste trabalho é fazer com que os investidores pessoa física possam estudar ações presentes no mercado de capitais brasileiro de maneira direcionada, com base em seu próprio perfil de investidor, com suas respectivas aversões a risco, tolerâncias a volatilidade, entre outros. Para isso, serão utilizadas técnicas de agrupamento presentes nas literaturas da inteligência artificial, para que possam selecionar e agrupar por afinidade determinados grupos de ações, para que investidores menos experientes possam mergulhar em estudos previamente analisados, e saibam o tipo de informação e ativos que estão lidando. Não se trata de recomendação de compra ou de venda de nenhum ativo, mas sim da pré-seleção de ações de acordo com determinados perfis de investidor, para que se possa haver um estudo guiado.

Palavras-chave—*perfil de investidor, ações, mercado de capitais, agrupamento, inteligência artificial*

I. INTRODUÇÃO

Atualmente, a bolsa de valores brasileira conta com mais de 400 empresas de capital aberto listadas e sendo negociadas todos os dias. Entretanto, para pequenos investidores pessoa jurídica e todos os investidores pessoa física, muitas dessas empresas e o conhecimento sobre elas fica inacessível para estudo, visto que apenas uma pequena parcela de todas as oportunidades são divulgadas pelas mídias especializadas e pelos grandes índices do mercado de capitais, como Ibovespa, entre outros. Com base nisso, o objetivo principal deste projeto é compreender os principais objetivos do investidor, entendendo seus objetivos a médio e longo prazo, e devolvendo para ele um conjunto de ações mais encaixadas com suas necessidades, para que o usuário possa estudá-las e definir o investimento ou não.

O mercado financeiro brasileiro movimentou R\$7,04 trilhões em 2021, e é um setor da economia que cresce constantemente ano após ano. Em fevereiro de 2022, eram 4,2 milhões de CPFs na bolsa de valores, ante 3,4 milhões de CPFs em novembro de 2021, o que mostra que o mercado de capitais brasileiro está em um crescimento exponencial.

Uma dificuldade fortemente enfrentada por todas as pessoas físicas novas na renda variável da bolsa de valores é, além de saber se um ativo é bom ou não para investimento, saber qual ativo exatamente estudar. Atualmente, existem poucos índices de qualidade listados na B3 que são conhecidos pelas pessoas, e apenas as ações contidas neles são comentadas e divulgadas por casas de análises ou mídia especializada.

Em 18/04/2022, havia 549 ações listadas na bolsa, referentes a 410 empresas (entre ações ordinárias e preferenciais), enquanto apenas 96 ações fazem parte do Ibovespa e chegam aos olhares de investidores menos experientes.

Tal falta de divulgação e conhecimento faz com que muitas pessoas, apenas por não saber que determinadas ações existem, deixem de realizar bons estudos e descobrir investimentos que se adequem a seus perfis.

Um exemplo que complementa a tese é que as ações de Magazine Luiza (MGLU3) valorizaram 382% entre a abertura de capital da empresa e a entrada dela no índice Bovespa, em 02/01/2018, valorização que foi perdida pelos investidores que estudam e pensam em investir apenas em ações do Ibovespa.

Com este projeto, será possível realizar testes de agrupamento, considerando os diversos perfis de investidor, de acordo com determinadas teses pré-estabelecidas. Um exemplo válido para se levar em consideração é o endividamento de uma empresa; é senso comum dentro do mercado financeiro que investidores mais conservadores optam por menores índices de endividamento em empresas às quais se propõem a investir, visto que estas estão menos suscetíveis à riscos de crédito e mais seguras diante de crises econômicas que possam

vir a enfrentar. Com a definição desses parâmetros e com uma ampla base de dados (indicadores financeiros de cada empresa, além de preços de fechamento diários dos últimos anos), é possível formar uma base de dados extremamente sólida para analisar de maneira minuciosa, e definir grupos de ações para grupos de investidores.

Portanto, torna-se interessante estudar o tema com o objetivo de ajudar o investidor médio, que não tem capital suficiente para pagar grandes analistas e nem as altas taxas propostas por gestoras financeiras que dedicam 100% de seu tempo a isso; permitindo, assim, que estes possam fazer suas próprias análises, entendendo previamente qual o seu perfil de investidor, e assim podendo buscar dentro de um “pote” de ações, quais ele pode estudar, sabendo que já está fazendo um estudo direcionado.

II. CONTEXTUALIZAÇÃO

Os conceitos de perfil de investidor servem para guiar os objetivos dos investidores sobre alocamento de seu capital, além de definir os interesses destes referente aos produtos disponibilizados por bancos e corretoras. Além disso, é muito útil para definir, principalmente, o apetite a risco que cada investidor possui. O perfil é definido de acordo com questionário realizado por bancos e corretoras no momento da abertura da conta, e pode ser refeito de acordo com mudanças de objetivo ou preferências de quem possui a conta. O perfil de investidor serve, também, como uma proteção por parte da CVM (Comissão de Valor Mobiliários) e das corretoras para “proteger” investidores inexperientes que, por não conhecerem corretamente os produtos e a dinâmica do mercado, podem se expor a riscos desnecessários, e futuramente alegar que não foram alertados pelos agentes do mercado.

Segundo Rambo (2014) [II], Almeida e Cunha (2017)[III], os perfis de investidores podem ser divididos da seguinte maneira:

Conservador: é o investidor que prioriza a segurança em suas aplicações, opta por investimentos com baixa oscilação e risco, ou seja, está disposto a abrir mão de rentabilidade em troca de mais segurança e liquidez, pois seu objetivo principal é a preservação de patrimônio, e prefere investimentos com maior liquidez, ou seja, que podem ser resgatados e transformados em dinheiro com mais rapidez.

Moderado: encontra-se entre os investidores conservadores e arrojados, procura equilibrar rentabilidade versus risco, admite perdas em curto prazo até certo limite, por não ter necessidade de retirar o dinheiro no curto prazo, permite maiores riscos e oscilações.

Arrojado: é o investidor que entende que as perdas a curto prazo são momentâneas e necessárias para aproveitar lucros mais altos a longo prazo, opta por assumir mais riscos em troca de maior rentabilidade e aceita maior volatilidade em sua carteira.

Entretanto, fora do que se considera o “padrão” dentro do mercado financeiro, é possível investir em ações de maneira mais conservadora ou moderada, de maneira que se obtenha segurança patrimonial de alta qualidade, mesmo sem correr muitos riscos, uma vez que existem diversas empresas no

mercado acionário brasileiro que nunca deram prejuízo em sua história, empresas que têm baixíssima volatilidade de sua cotação e, ainda, empresas que distribuem quantidades generosas de seus lucros através de dividendos.

Tais fatos vão na direção completamente oposta ao que é vendido sobre perfis de investidor, mostrando de maneira objetiva que o investimento em renda variável não é composto apenas de volatilidades gigantes, grandes ganhos ou perda de capital, acompanhamento dos gráficos 24 horas por dia e aqueles cenários que lembram os cassinos americanos, mas sim de pessoas que querem investir em negócios sérios e formar patrimônios sólidos ao longo da vida.

Um exemplo prático dos dias de hoje que corrobora com a tese do projeto é o cenário econômico da Argentina. Segundo publicado pelo portal de notícias da B3 [1], no dia 15/05/2023, o Banco Central Argentino elevou sua taxa de juros básica para 97% ao ano, ou seja, um investimento conservador realizado por um argentino, hoje, rende 97% no período de um ano. Entretanto, um ponto importantíssimo na discussão é a inflação presente no país, que no mês de abril deste ano, atingiu 108,8% ao ano. Em contrapartida, o índice geral da bolsa de valores argentina, o índice Merval, se valorizou em 264,71% entre 15 de maio de 2022 e 2023. Ou seja, qualquer investidor argentino, por mais conservador e com maior aversão a risco que possua, tem que investir seu dinheiro na bolsa de valores, se não quiser perder dinheiro para a inflação; isso se dá pois a bolsa de valores negocia ações de empresas reais da economia, que podem repassar seus preços para consumidores, e são capazes de reajustar seus faturamentos e, consequentemente, sua cotação, acima da inflação.

III. TRABALHOS RELACIONADOS

Após pesquisa em diretórios acadêmicos online, foi possível encontrar alguns trabalhos relacionados ao assunto, ou que tratem de finanças, ações, métricas de agrupamento, ou que unam todos esses pontos. Eles são enumerados e explicados a seguir e, além disso, são melhor explicados ao longo do artigo, de acordo com a demanda do texto:

- I. M. A. Ferreira, R. Grossi, C. Leal AND S. Silveira, “Perfil de investidores e formação de portfólios: contribuições a partir da análise multivariada de dados”, Revista de Negócios, 2008 [online], disponível em: https://www.researchgate.net/publication/287110663_PERFIL_DE_INVESTIDORES_E_FORMACAO_DE_PORTFOLIOS_CONTRIBUICOES_A_PARTIR_DA_ANALISE_MULTIVARIADA_DE_DADOS
- II. RAMBO, A. C. O perfil do investidor e melhores investimentos: da teoria à prática do mercado brasileiro. Monografia (Bacharel em Ciências Econômicas), Universidade Federal de Santa Catarina, Florianópolis, 2014.
- III. ALMEIDA, A. L. F.; CUNHA, D. P. A. Estudo do mercado brasileiro de renda fixa e o perfil do investidor brasileiro. Monografia (Bacharel em Engenharia de Produção), Escola Politécnica – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.
- IV. VARGAS, L. F.; SOARES, G. L.; COSTA JR. P. P. Ferramenta de Inteligência Artificial para Seleção de Carteira de Ações no Setor Elétrico. Programa de Pós-graduação em Engenharia Elétrica PPGEE. PUC-MG, Belo Horizonte, 2015.

- V. OLIVEIRA, L.L. Proposta de uma carteira de investimentos em tesouro direto para o perfil de investidor de uma instituição pública de ensino. Universidade Federal de Ouro Preto, João Monlevade, 2020.
- VI. DE CASTRO, L. N.V. Análise do Perfil do Investidor: Desenvolvimento e Validação de Questionário Padrão. Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2012.
- VII. RODELLA, V. G. Estudo de caso: aplicação de machine learning para a previsão de tendências das ações das bolsas de valores brasileira e norte americana. Instituto de Ciência e Tecnologia, Universidade Estadual Paulista, Sorocaba, 2023.
- VIII. SILVA, W.V et. Al. Finanças comportamentais: análise do perfil comportamental do investidor e do propenso investidor. Revista Eletrônica da Ciência Administrativa. Vol. 7 N°2, 2008.
- IX. XXVII SIMEP. CLUSTERIZAÇÃO DE ATIVOS: UMA ALTERNATIVA PARA AUXILIAR NAS DECISÕES FINANCEIRAS[Online] Disponível em: https://www.researchgate.net/profile/Daiane-Santos-5/publication/346299242_CLUSTERIZACAO_DE_ATIVOS_UMA_ALTERNATIVA_PARA_AUXILIAR_NAS_DECISoes_FINANCEIRAS/links/5fbd9e73458515b79769f8d2/CLUSTERIZACAO-DE-ATIVOS-UMA-ALTERNATIVA-PARA-AUXILIAR-NAS-DECISoes-FINANCEIRAS.pdf
- X. TORRES, R. A. Aplicação de Métodos de Clusterização em um estudo sobre o Mercado Acionário Brasileiro. Departamento de Matemática da PUC-Rio. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

No estudo I, os autores utilizam atributos mensuráveis do comportamento de ações da bolsa brasileira para simular a criação de portfólios de acordo com perfis de investidor. Para tal, eles usam a teoria do Portfólio, e fazem uma análise multivariada dos dados para maximizar a utilidade esperada de um investidor. No estudo, eles conseguem associar variações de resultados de acordo com os respectivos perfis.

No estudo II, a autora faz uma análise dos perfis de investidores presentes no mercado financeiro e faz uma listagem de todos os investimentos presentes e quais são mais adequados para cada perfil. Para tal, ela faz uma contextualização teórica de todos os componentes do mercado, as taxas que o regem e as opções de investimentos fornecidas por bancos e corretoras. Para isso, ela utiliza o conceito de perfil de investidor que foi apresentado no item II deste artigo, sendo essa a contribuição para nosso trabalho.

No estudo III, os autores fazem uma análise do mercado de renda fixa brasileiro, e fazem um paralelo com os perfis de investidor presentes no mercado. A contribuição deste artigo para nosso projeto é semelhante à presente no estudo II, pois nos forneceu o conceito de perfil de investidores utilizados por bancos e corretoras. Tal conceito já foi explicado no item II deste artigo.

O estudo IV mostra o desenvolvimento de uma ferramenta de inteligência artificial que objetiva a seleção de ações do setor elétrico para os diversos perfis de investidores do mercado financeiro. Para tal, a aplicação usa indicadores macroeconômicos para tentar prever a rentabilidade de tais investimentos.

O Estudo V consiste na proposta de criação de uma carteira de investimentos em tesouro direto com base no perfil de investidor de uma instituição de ensino específica. Este artigo contribui para nosso projeto pois utiliza a metodologia API (Análise de perfil de investidor), que é importante para definir os melhores indicadores que se adequem às vontades dos investidores.

Em concordância com o estudo V, o estudo VI realiza uma validação dos questionários de perfil de investidor apresentados por bancos e corretoras para definir perfis de risco, liquidez e volatilidade dentro dos investimentos. Assim como o estudo V, entender metodologia de perfil do investidor nos auxilia para compreender quais indicadores econômicos das ações devem ser mais considerados nas tarefas de agrupamento.

No estudo VII, o autor busca a utilização de técnicas de machine learning para prever tendências de ativos na bolsa de valores brasileira e norte-americana. Como será citado mais a frente neste artigo, o autor deste estudo utiliza indicadores do próprio ativo para realizar suas previsões, enquanto outros utilizam indicadores macroeconômicos. Esta análise será utilizada como repertório na definição dos indicadores a serem considerados para a separação em clusters.

No estudo VIII, grupo de pesquisadores utiliza de um questionário com amostragem em sua instituição de ensino, no qual se obtém maiores informações sobre os perfis de investidores e de propensos investidores, a fim de evidenciar irracionalidades dos investidores, para ordená-los em grupos de investidores, considerando seus principais objetivos.

O estudo IX trata sobre a clusterização de ativos do mercado financeiro para auxiliar na tomada de decisões. Este artigo será de grande utilidade para o projeto por seu caráter técnico, uma vez que ele explora a divisão de grupos de ações em clusters, e sua posterior análise por indicadores financeiros, como rentabilidade e volatilidade. Ele serve como base técnica para este projeto, por realizar a respectiva modelagem e separação dos ativos em grupos de maneira semelhante ao proposto neste trabalho.

Em linha com o estudo IX, o estudo X utiliza de métodos de clusterização para analisar se os valores de fechamento de ações do mesmo setor da economia se apresentam semelhantes, ou seja, se empresas do mesmo setor possuem o mesmo grau de rentabilidade em suas cotações. Com a separação por clusters e 3 diferentes métodos de análise de correlação, foi possível encontrar apenas 1 grupo em que haja a correlação da tese. Este artigo é de utilidade para o projeto por seu fator técnico, uma vez que este projeto também objetiva a separação por cluster e análise de indicadores financeiros.

IV. OBJETIVO

Os objetivos do projeto, portanto, são permitir a futuros investidores que possam respeitar seus perfis, sem que isso limite quais produtos eles possam acessar; sendo assim, permitindo que os mais diversos perfis de investidor tenham acesso e se exponham à bolsa de valores de acordo com seus objetivos. Assim, isso permitirá a todos a sofisticação e

profissionalização de seus investimentos, quebrando os mitos existentes sobre o mercado acionário.

V. METODOLOGIA EXPERIMENTAL

A metodologia utilizada no projeto em questão será a de técnicas de agrupamento, partindo de uma base de dados fornecida pelo site Status Invest. O Status Invest é uma das maiores plataformas de dados financeiros do Brasil. Desenvolvida por analistas do mercado financeiro em parceria com a casa de análise Suno Research, o site fornece todas as informações sobre cada ativo listado na bolsa de valores brasileira, e alguns listados em bolsas internacionais. Na plataforma, você encontra desde o CNPJ de uma empresa ou fundo, como relação de preço/valor patrimonial, indicadores de crescimento de faturamento e lucros, além dos balanços financeiros completos, tanto de ações, fundos imobiliários e ETFs. Neste projeto, as informações serão extraídas do Status Invest através de um arquivo excel. Além disso, será utilizado o Yahoo Finance, que é uma ferramenta financeira, bem como um braço de finanças do conhecido veículo de informações Yahoo. Nela, é possível ver as últimas notícias do mercado financeiro brasileiro, americano e mundial, além de cotações da bolsa de valores, notícias sobre a governança das companhias e outras atualizações voltadas para o assunto de finanças.

Por ter as cotações de todos os ativos em tempo real, ela serve como banco de dados que será consumido pela biblioteca Pandas, de onde serão extraídas as informações de cotações dos últimos anos a fim de se realizar diversas explorações e análises, como a volatilidade de cada ativo. O projeto será desenvolvido em Python e utilizará algumas das bibliotecas voltadas para inteligência artificial, análise exploratória e visualização de dados, como por exemplo Scikit-learn, Tensorflow, Matplotlib, Seaborn, Pandas, Numpy, Apyori entre outros.

Tendo disponível para estudo a base de dados, primeiramente será realizado um processo de pré processamento e preparação dos dados. Eventualmente os dados serão tratados efetuando a limpeza, normalização ou até mesmo a redução da dimensionalidade do problema com a finalidade de garantir a integridade e consistência dos dados para análise.

Posteriormente, antes de se aplicar os algoritmos de aprendizado de máquina, os dados passarão pela etapa de visualização por meio de histogramas, gráficos, diagramas de dispersão, correlação, distribuição estatística entre outros com o intuito de identificar a similaridade e relevância dos mesmos para o problema em questão, sendo assim possível selecionar ou até mesmo, extrair novas features (características) relevantes para análise. Tendo as features selecionadas devidamente, o próximo passo do processo é a modelagem, onde serão aplicadas divisões da base de dados para teste e treino, realização das tarefas de agrupamento de dados, no fim desse processo, teremos um modelo treinado e prestes a ser realmente testado. Na etapa de teste, o desempenho do modelo será averiguado por meio de métricas de avaliação como por exemplo, as medidas de precisão, recall, erro quadrático médio (RMSE) e assim por diante.

Com base nos resultados obtidos no processo de validação, o modelo poderá ser ajustado, melhorando a seleção de hiper parâmetros, eventuais mudanças de algoritmo e aplicação de técnicas de regularização.

Por fim, segue a validação final, etapa onde após o modelo ser otimizado, é importante realizar um teste final com novos dados ou por meio da técnica de validação cruzada.

Um exemplo bibliográfico de aplicação semelhante é o artigo I, citado no item 3 – Trabalhos Relacionados, no qual os autores utilizam atributos mensuráveis do comportamento de ações do mercado de capitais brasileiro para simular a criação de portfólios de acordo com perfis de investidores. Para tal, eles usam a teoria do Portfólio, e fazem uma análise multivariada dos dados para maximizar a utilidade esperada de um investidor. No estudo, eles conseguem associar variações de resultados de acordo com os respectivos perfis.

Embora ainda não tenham sido escolhidos os indicadores econômicos que serão utilizados na seleção dos ativos por perfil de investidor, serão utilizados como parâmetro os estudos de Ferreira, Grossi, Leal e Silveira (2008); Vargas, Soares e Costa Jr. (2015) [IV] e Rodella (2023), que utilizaram tanto indicadores macroeconômicos, como PIB, IPCA e SELIC, como indicadores dos próprios ativos, como liquidez, volatilidade e distribuição de dividendos, além das diferenças entre análise gráfica e fundamentalista. Vale ressaltar que, neste ponto do projeto, os autores estão inclinados a utilizar informações e indicadores dos próprios ativos, ante a indicadores macroeconômicos, embora isso ainda não tenha sido decidido.

Além disso, serão utilizados como base os conceitos aplicados e estudados por Oliveira (2020) e De Castro (2012) no que se refere à Análise de Perfil de Investidor (API), que consiste no questionário desenvolvido e utilizado para determinação do perfil de investidor daqueles que o respondem. Os parâmetros e objetivos do investidor, que são utilizados para chegar a uma resposta final, serão levados em consideração neste projeto.

A. Base de dados

A base de dados utilizada nesse trabalho consiste em dados discretos contidos em um documento em Excel com as informações de todas as ações e empresas brasileiras listadas em bolsa de valores. A tabela foi retirada do site Status Invest, que já foi contextualizado anteriormente neste artigo, e é uma das bases de dados mais respeitada entre os agentes do mercado financeiro. A base conta com 621 linhas, cada uma representando uma ação da bolsa brasileira, e com 30 colunas, que representam os diversos indicadores que podem ser utilizados para a análise de uma empresa. Os indicadores listados na tabela são os seguintes (o símbolo “/” significa “dividido por”):

Ticker (nome), preço, Dividend Yield, Preço/lucro, Preço/Valor patrimonial, Preço/ativos, margem bruta, margem Ebit, margem líquida, preço/ebit, EV/ebit, dívida líquida/ebit, Dívida líquida/patrimônio, PSR (preço/receita operacional), Preço/capital de giro, preço/ativo circulante líquido, liquidez corrente, ROE (retorno sobre participação),

ROIC (retorno sobre investimento), ROA (retorno sobre ativos), patrimônio/ativos, passivo/ativos, giro ativos, CAGR receita 5 anos (crescimento médio da receita nos últimos 5 anos), CAGR lucro 5 anos (crescimento médio do lucro nos últimos 5 anos), liquidez média diária, VPA (valor patrimonial por ação), LPA (lucro por ação), PEG Ratio e valor total de mercado. A lista com as ações pode conter mais de uma ação de uma empresa, pois há 3 tipos de ações (ordinárias, preferenciais e units), os 3 tipos serão considerados para a análise, por representarem aspectos diferentes dentro do mundo dos investimentos. Ações ordinárias possuem direito a voto no conselho administrativo de uma empresa, ações preferenciais possuem prioridade na distribuição de dividendos e units são “pacotes” de uma certa quantidade de ações ordinárias e preferenciais, variando de empresa para empresa. Os dados correspondem aos valores dos indicadores no dia 12 de junho de 2023.

Em seguida, foi realizada a limpeza nos itens da base de dados, para que os algoritmos de agrupamento e validação utilizados possam ser aplicados corretamente. Assim, todos os dados que constam como NULL foram transformados em 0. Os valores de NULL não foram removidos, por simbolizarem que não há dados suficientes para calcular estes indicadores para tais empresas, o que não invalida a empresa como uma possibilidade de investimento. Um exemplo são os valores de DY que constam como NULL. DY é o cálculo de Dividend Yield de uma empresa, ou seja, quanto de lucro que foi distribuído ao acionista dividido pelo valor de mercado da empresa. Se o valor de Dividend Yield é nulo, não quer dizer que haja algo de errado com a empresa, mas apenas que ela não distribuiu lucros aos seus acionistas, e prefere utilizar o dinheiro para reinvestir no negócio e focar no aumento de sua base de clientes e faturamento. Empresas que não distribuem dividendos costumam ser mais atrativas para investidores mais arrojados, enquanto empresas que distribuem boa parte de seus lucros são interessantes para investidores mais moderados, pois são empresas mais consolidadas que focam em manter sua presença no setor, e tomam decisões menos arriscadas para crescer.

Uma exceção foi no caso dos indicadores de CAGR de receita e de lucro nos últimos 5 anos. O CAGR representa a média de crescimento, tanto de receita como de lucro nos últimos 5 anos. Nos casos destes indicadores, estarem nulos, não quer dizer que não houve crescimento, pois isto seria simbolizado através de números, mas sim que não há dados suficientes para se calcular tais fatores nos últimos 5 anos. Isso se dá majoritariamente em empresas que são listadas na bolsa de valores a menos de 5 anos, ou seja, suas receitas e lucros não são públicas há tanto tempo, ou seja, simplesmente não há dados para se calcular este indicador. Por isso, colocamos o valor numérico 0.01 nesses casos, pois por se tratar de empresas com receitas e lucros milionários, sendo absurdo ter um crescimento de 0.01 ponto percentual na média do período. Assim, 0.01 serve como um indicador numérico de que não há dados no período.

Obs: Empresas que têm quedas nas suas receitas e/ou lucros nos últimos 5 anos possuem um valor de CAGR negativo.

Além disso, foi realizada a limpeza de dados que de fato não contribuirão para o posterior treinamento e teste dos dados, como foi o caso de 7 ações sem informação de valor de mercado, ou seja, estão contidas no documento, mas não possuem mais valor, seja por quê pararam de ser negociadas, ou não há dados suficientes para o fornecedor dos dados gerar algum valor. Além disso, foram retiradas 134 ações sem informações referentes à liquidez média diária, pois ao não possuírem esta informação, quer dizer que dado referente à negociações nos últimos pregões. Isso quer dizer que a ação não está mais em negociação, ou seja, foi comprada por outra empresa ou recomprou suas ações no mercado.

Além disso, foi adicionado ao código a análise estatística dos dados como cálculo da correlação e a função describe do pandas, que calcula média, desvio padrão, separação por quartis de todas as colunas do dado, com o objetivo de ter um melhor repertório ao propor os testes a serem realizados.

B. Protocolo de validação

O protocolo de validação que será utilizado no projeto será o Silhouette score, que faz uma análise de compactação dentro de um cluster e o nível de separação entre os clusters. Os valores do Silhouette score variam de -1 a 1, sendo que 1 indica bons agrupamentos, -1 indica agrupamentos completamente errados e 0 que há clusters sobrepostos. Os dados utilizados para calcular o Silhouette score são compostos da distância média intra-cluster, e a distância média para o cluster mais próximo (inter-cluster). Quanto menor a distância média para os itens intra-cluster, mais agrupado é o cluster, o que indica melhor qualidade do agrupamento. Quanto maior a distância média no valor inter-cluster, mais separados são os clusters entre si, o que também indica um bom agrupamento.

O cálculo do Silhouette Score é feito da seguinte maneira: $\text{Silhouette_score} = (b - a) / \max(a, b)$. Onde (a) representa a distância intra-cluster e (b) a distância inter-cluster. A partir disso se chega ao resultado final, onde quanto mais próximo de 1, melhor o agrupamento realizado.

O cálculo do Silhouette score será realizado para todos os testes deste projeto, tanto testes em que há variação nos parâmetros utilizados, como nos testes com variação na quantidade de clusters utilizados no agrupamento.

A seguir, segue um diagrama de blocos com o pipeline experimental do projeto, com todas as etapas e suas respectivas ordens de realização:

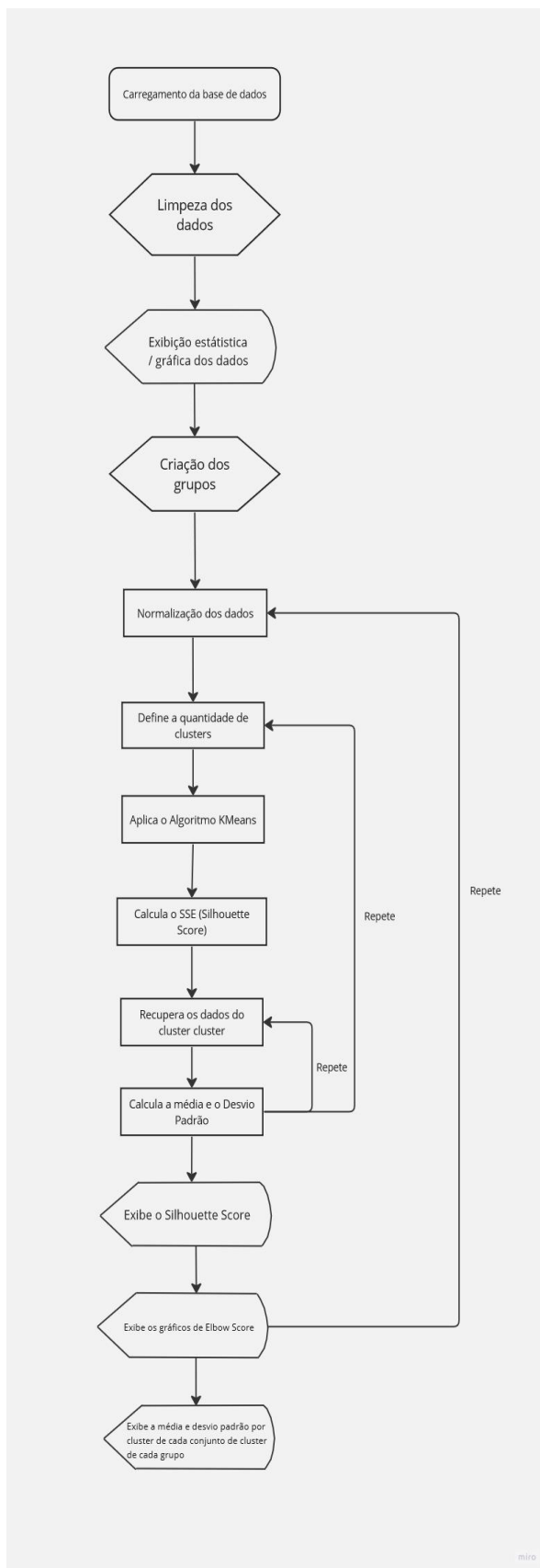


Imagem 1: Diagrama de blocos com pipeline experimental do projeto
Fonte: Autores

VI. CONCEITOS IMPORTANTES

O principal modelo de IA utilizado neste projeto será o modelo de clusterização K-means, no qual os dados são divididos por K clusters. Com isso, os dados serão separados em grupos de ações de acordo com as suas afinidades com base em indicadores pré-definidos. Além disso, para cada valor de k clusters, será realizada a validação de Silhouette Score para verificar quão bom é o agrupamento feito através do algoritmo. Além disso, escolhemos o método de agrupamento K-medoids para posterior implementação e comparação no próximo relatório com o método atualmente utilizado. O mesmo vale para o protocolo de validação, no qual foi escolhido o Elbow Score, o qual também será implementado e rodado para comparar com a comparação de Silhouette Score que está sendo feita. O objetivo destas comparações é aumentar a validação dentro do projeto, para se encontrar o método mais preciso para o agrupamento proposto. O objetivo é encontrar uma quantidade de clusters para o qual faça sentido a separação entre os grupos, além disso, serão realizados testes para os 5 seguintes grupos de indicadores.

0: DY, dívida líquida / EBIT, dívida líquida / patrimônio, CAGR receita 5 anos, CAGR lucro 5 anos

1: DY, CAGR lucro 5 anos, liquidez média diária, P/L, margem líquida

2: liquidez média diária, Preço / Lucro, margem líquida, ROIC, VPA

3: liquidez média diária, DY, margem líquida, dívida líquida / EBIT, VPA

4: CAGR receita 5 anos, margem líquida, VPA, P/ATIVOS, dívida líquida / EBIT

Os grupos de indicadores foram escolhidos por representar as informações relevantes que são consideradas ao se avaliar ativos no mercado financeiro. Elas contemplam informações das empresas, como estrutura de capital (indicadores de dívida, PVA), indicadores de capacidade de geração de caixa (CAGR receita, CAGR lucro e Margem líquida), e geração de valor para o acionista (DY). Além disso, estes indicadores em questão são ótimos indicadores para medição de risco por parte das empresas, como explicado anteriormente no artigo. Dívidas maiores indicam maior risco, liquidez média diária menor indica maior risco, e vice-versa. Além disso, por questões de teste, indicadores que aparecem em vários grupos são retirados de outro, para testar o seu impacto na qualidade do agrupamento. Um exemplo disso é o DY, que está presente em 3 grupos e foi deixado de fora nos grupos 2 e 4; estes grupos, por sua vez, apresentaram os melhores resultados de Silhouette Score. A separação dos respectivos indicadores entre os grupos não seguiu nenhuma estratégia específica, além da retirada de alguns em certos grupos.

Para cada um dos grupos, foi realizada a clusterização K means, e para cada valor de k clusters, foi calculado o Silhouette Score. Inicialmente, os valores de K variaram de 2 a 10 clusters por grupo, entretanto, foi possível perceber que separações a partir de 5 clusters encontravam um valor de Silhouette Score muito baixo, e por isso optou-se por refazer os testes para valores de 2 a 5 clusters por grupo testado.

O passo a passo do projeto dar-se-á da seguinte maneira: tal qual feito pelos autores do trabalho relacionado I, separar o grupo de ações presente na base de dados em clusters, realizando testes para diferentes indicadores e diferentes quantidades de clusters para cada um deles. Ao mesmo tempo em que há a separação desses clusters, fazer a verificação do Silhouette Score. Em seguida, montar uma tabela com os valores da média e do desvio padrão de cada indicador para os 5 grupos testados, e para o melhor valor do Silhouette Score para o número de clusters. Por fim, na discussão de resultados, apresentar e atribuir os grupos aos perfis de investidores previamente explicados, concluindo, por fim, se é possível chegar a uma atribuição total dos grupos de acordo com os perfis, se é possível fazer isso de maneira parcial ou se não há relação nenhuma entre os resultados encontrados nos agrupamentos e os perfis apresentados.

Para que qualquer tipo de atribuição possa ser feita entre um indicador em específico e um perfil de investidor, é necessário entender quais valores de cada indicador mais se aproximam dos perfis de investidor, como será mostrado a seguir:

DY: Dividend Yield é a distribuição de lucro aos acionistas, ou seja, ao fechamento de um semestre, a empresa computa seu lucro e decide se o retém, ou se transfere para a conta do acionista. Empresas que distribuem bastantes dividendos são conhecidas por serem mais consolidadas, ou seja, menos arriscadas. Empresas menos consolidadas e que precisam gastar mais dinheiro para expansão tendem reter seus lucros para reinvestir e, portanto, são mais atraentes para investidores arrojados, com mais apetite a risco.

Dívida Líquida / EBIT: é o cálculo da dívida líquida de uma empresa dividido pelo seu lucro antes de impostos e juros (EBIT), este indicador avalia o endividamento de uma empresa, e indica quanto tempo seria necessário para pagar a dívida líquida de uma empresa considerando o EBIT atual. Empresas com este indicador mais alto são mais endividadas e, naturalmente, estão mais de acordo com o perfil de investidor arrojado, que assume mais riscos. Empresas endividadas possuem mais riscos relacionados à flutuação da taxa de juros.

Dívida Líquida / Patrimônio: é o cálculo da dívida líquida de uma empresa dividido por seu patrimônio. Semelhante ao indicador explicado acima, este também avalia o endividamento de uma empresa. Sendo que se o valor estiver abaixo de 1, a empresa pode, em casos extremos, se desfazer de todo o seu patrimônio para pagar a dívida, e sair no zero a zero. Também de acordo com explicado acima,

empresas com este valor maior possuem mais riscos, e estão mais de acordo com o perfil de investidor arrojado e vice-versa.

Liquidez Média Diária: é o valor referente à média de liquidez do ativo nos últimos 30 dias. A liquidez indica a rapidez com que se pode transformar determinado ativo em dinheiro. Uma liquidez menor indica mais demora para transformar um ativo em dinheiro, ou seja, em casos de emergência seria ideal poder ter o dinheiro na mão o mais rápido possível, pois não se pode pagar um hospital em ações, por exemplo. Assim, empresas e ativos no geral, que possuam uma liquidez média diária mais alta são menos arriscados, e estão mais de acordo com o perfil de investidor conservador, e vice-versa.

Preço / Lucro: é o cálculo do preço dividido pelo lucro, e indica quantas vezes o lucro vale a empresa. Quando o P/L de uma empresa é muito alto, encontra-se risco, pois pode ser um sinal de uma bolha, e é importante comparar este indicador com outros ativos de setores semelhantes. Investidores que investem em empresas com P/L muito alto fazem uma movimentação arriscada, e, portanto, têm mais apetite a risco, e vice-versa.

Margem Líquida: a margem líquida de uma empresa é a porcentagem de seu faturamento que se reverte em lucro líquido. Empresas e setores com maiores margens líquidas são mais seguros, e são mais resistentes a crises e mudanças no mercado. Empresas com intenção de crescimento mais acelerado muitas vezes diminuem suas margens de lucro em busca de competitividade no mercado, além disso, baixas margens são menos resistentes a mudanças repentinas no mercado. Assim, investidores com menor apetite a risco preferem empresas com maiores margens, e vice-versa.

ROIC: Retorno sobre capital investido é o cálculo da capacidade de uma empresa de gerar dinheiro sobre o capital investido nela. Naturalmente, uma empresa que gere mais dinheiro com menor investimento é uma melhor escolha do que o contrário.

VPA: O valor patrimonial por ação é o cálculo do valor patrimonial da empresa dividido pela quantidade de ações. Ele se assemelha ao cálculo de preço / valor patrimonial no sentido de que se a relação VPA é maior que 1, ou seja, se o preço da ação é maior que o valor patrimonial por ação, o valor da empresa se dá além de seus ativos físicos e financeiros, mas sim pela sua capacidade de gerar lucro ao longo do tempo. Um VPA maior que o valor da ação pode ser um investimento menos arriscado, ou até uma oportunidade, enquanto um VPA muito abaixo do valor da ação pode indicar uma euforia descabida perante a ação, sendo assim mais arriscado.

Preço / Ativos: é o cálculo do preço da ação / ativos por ação, o que equivale ao valor total da empresa dividido pelo valor dos ativos da empresa. É um indicador da saúde financeira da empresa, semelhante ao cálculo da dívida / patrimônio. Se o valor for acima de 1, quer dizer que o valor atrelado à empresa se dá além dos ativos físicos e financeiros que ela possui, mas também pela sua capacidade de gerar dinheiro. Uma empresa com o indicador preço/ativos baixo

pode indicar uma oportunidade, assim como esse indicador extremamente alto também pode indicar uma bolha na ação.

É de extrema importância ressaltar que nenhum indicador deve ser considerado sozinho na formulação de uma tese de investimentos, e é por esse motivo que eles estão colocados em grupos para seus respectivos agrupamentos.

Entrega Final

A entrega final consiste no código desenvolvido pelos autores do trabalho, que consistirá na importação dos dados, tratamento e limpeza dos dados, cálculos necessários para se chegar aos indicadores que serão utilizados (Ex: cálculos de volatilidade, valorização, volume de negociação). Além disso, serão atribuídas determinadas características aos respectivos ativos, para que se possa chegar às regras de agrupamento.

Além disso, o relatório final contará com a separação por grupo entre os perfis de investidor, e os níveis de volatilidade, crescimento, Dividend Yield, e outras informações que serão utilizadas para separar os perfis.

A entrega final consistirá na discussão e conclusão se os valores agrupados em cada grupo dos testes de agrupamento podem ser atribuídos aos seus respectivos perfis de investidores, como feito no trabalho relacionado I. Serão apresentados os números e será discutido se estes valores se aplicam às logicas de risco e os conceitos previamente apresentados para cada perfil de investidor. O resultado final será a atribuição de determinado cluster de ações à determinado perfil de investidores.

Após a comparação dos modelos de IA, das validações, das medidas de avaliação e dos testes de diferenças estatísticas adotados, e ter encontrado o grupo de indicadores e a quantidade ideal de clusters no qual foi feito o agrupamento, averiguaremos se é possível atribuí-los aos perfis, chegando a uma resposta final.

VII. RESULTADOS

Nesta etapa do projeto já foram realizados diversos testes considerando os 5 grupos de indicadores, com os valores de k-means variando de 2 a 5 grupos. Assim, obtivemos os seguintes valores de Silhouette Score para cada grupo e para cada valor de k-means.

GROUP =	K =	SILHOUETTE SCORE
0	2	0.7149315031713313
0	3	0.3126986750491167
0	4	0.4243974067606472
1	2	0.9116948060008507

1	3	0.501250649614624
1	4	0.5007914153239582
2	2	0.9652818361308114
2	3	0.90541173938754
2	4	0.9138767347774824
3	2	0.8421717097282432
3	3	0.7692745121142202
3	4	0.775994995158199
4	2	0.9327107423805142
4	3	0.877050813662048
4	4	0.8605219263446486

Tabela 1: Silhouette Score para os grupos e valores de k-clusters
Fonte: Autores

A partir de uma primeira análise nos valores de validação, é possível compreender que o grupo 0 contém os piores valores, com 0,71 para k=2; 0,31 para k=3 e 0,42 para k=4, e consequentemente possuem o agrupamento menos preciso. Em contrapartida, é possível averiguar que o grupo 2 possui os melhores valores de validação, com aproximadamente 0,96 para k=2; 0,90 para k=3 e 0,91 para k=4 clusters. Para o segundo melhor grupo, o grupo 4 apresentou ótimos valores: (0,93) para k=2, (0,87) para k=3, mas com abrupta queda (0,86) quando k=4; e o grupo 3 apresentou valores sólidos, porém mais baixos, com 0,84 para k=2, 0,76 para k=3 e 0,77 para k=4. Assim, os melhores grupos para o cálculo são os grupos 2 e 4

A partir de agora realizaremos a avaliação da média dos valores encontrados de acordo com os grupos:

Para o grupo 2 com 2 cluster, no qual encontramos 0,96 no Silhouette Score:

QTD_Cluster	N_cluster	Indicador	Valor
2	1	Liquidez Média Diária	4.155830^7
2	2	Liquidez Média Diária	0.000000
2	1	Preço / Lucro	39.542381
2	2	Preço / Lucro	-78669.16
2	1	Margem Líquida	131.49
2	2	Margem Líquida	0.000000

2	1	ROIC	17.346149
2	2	ROIC	0.000000
2	1	VPA	22.9494
2	2	VPA	0.0000

Tabela 2: Resultados média de valores para grupo 2 com 2 clusters
Fonte: Autores

Considerando a divisão entre 2 clusters do protocolo de agrupamento, é plausível atribuir o cluster 1 a um perfil de investimento conservador, visto que há alta liquidez média diária, altíssima margem líquida, um alto retorno sobre capital investido e alto valor patrimonial por ação. Embora o valor de preço/lucro esteja alto e traga certo risco, ao se analisar junto dos outros atributos é perfeitamente plausível entender este perfil como conservador.

Em contrapartida, o cluster 2 pode ser atribuído a um perfil extremamente arrojado, com pouca liquidez, baixo valor patrimonial e sem retorno sobre capital investido até o momento. Para um investidor investir nesta empresa, deve conter algum tipo de informação externa, como conhecimento do setor ou de seus fundadores. Investimentos com esse tipo de embasamento são comuns à investidores arrojados.

Agora, para o grupo 4 com 2 cluster, no qual encontramos 0,93 no Silhouette Score:

QTD_Cluster	N_cluster	Indicador	Valor
2	1	CAGR Receita 5 anos	12.876667
2	2	CAGR Receita 5 anos	0.010000
2	1	Margem líquida	131.497308
2	2	Margem líquida	0.000000
2	1	VPA	22.9494
2	2	VPA	0.000000
2	1	P/ATIVOS	1.19
2	2	P/ATIVOS	33255.410000
2	1	dívida líquida / EBIT	3.316439
2	2	dívida líquida / EBIT	0.000000

Tabela 3: Resultados média de valores para grupo 4 com 2 clusters
Fonte: Autores

Na análise do grupo 4, também fica clara a separação entre os clusters, uma vez que o primeiro cluster se encaixa com as características de um investidor conservador com CAGR de Receita de 5 anos em 12%, margem líquida altíssima, alto valor patrimonial, preço /ativos saudável e

dívida líquida/ EBIT saudável em 3.31. No cluster 2, na mesma toada do exemplo do grupo 2, o valor de CAGR indica uma empresa nova, além de um preço/ativos extremamente inflado e margem líquida baixa.

A separação por cluster com base nos indicadores dos clusters 2 e 4 pode ser considerada satisfatória, uma vez que combina bem os diversos indicadores considerados e pode ser atribuído a 2 dos três grupos de investidores.

Importante ressaltar que para o cluster 2 na tabela 2, a maior parte dos valores médios dos indicadores é 0, mas isso não necessariamente representa um erro ou algo do tipo, mas sim o que o valor destes indicadores é 0, o que tem suas próprias decorrências no sentido de risco e qualidade de determinado ativo.

Em seguida, realizaremos a análise dos mesmos grupos 2 e 4, mas desta vez para os grupos de 3 clusters, para verificar se o algoritmo separa de maneira correta os valores e estes podem ser coerentemente atribuídos aos 3 perfis de investidor previamente explicados neste artigo (conservador, moderado e arrojado).

Para o grupo 2 com 3 clusters, no qual encontramos 0,89 no Silhouette Score:

QTD_Cluster	N_cluster	Indicador	Valor
3	1	Liquidez Média Diária	4.19^7
3	2	Liquidez Média Diária	1.05^5
3	3	Liquidez Média Diária	0.00
3	1	Preço / Lucro	39.96
3	2	Preço / Lucro	-1.326
3	3	Preço / Lucro	-78669.16
3	1	Margem Líquida	119.44
3	2	Margem Líquida	1284.12
3	3	Margem Líquida	0.000000
3	1	ROIC	11.13
3	2	ROIC	611.13
3	3	ROIC	0.00
3	1	VPA	24.62
3	2	VPA	-137.0
3	3	VPA	0.00

Tabela 4: Resultados média de valores para grupo 2 com 3 clusters
Fonte: Autores

Avaliando a clusterização do grupo 2 com k= 3 clusters, é possível perceber que, o grupo 1 consta com uma alta liquidez média diária, um alto preço/lucro, alta margem líquida, um retorno sobre capital investido não tão alto, mas considerado saudável, e um alto valor patrimonial por ação, mas considerado saudável se comparado com o valor de preço/lucro. O grupo 2 apresentou liquidez média diária não

tão alta quando o grupo 1, mas também saudável, um valor preço lucro negativo, o que apresenta mais risco, margem líquida altíssima, retorno sobre capital investido altíssimo e um valor patrimonial por ação negativo. O grupo 3, por sua vez, apresentou liquidez média diária de 0, preço/lucro extremamente negativo, margem líquida de 0, retorno sobre capital investido de 0 e valor patrimonial de 0. Analisando estes valores, fica claro inferir que o grupo 3 é mais alinhado com o perfil arrojado de investidor, pois possui indicadores adversos e que possuem alto risco, com baixo valor patrimonial, sem retorno sobre capital investido até o momento e pouca liquidez, para um investidor investir neste grupo de empresas, deve conter algum tipo de informação externa, como conhecimento do setor ou de seus fundadores. Investimentos com esse tipo de embasamento são comuns à investidores arrojados.

Ao analisar os clusters 1 e 2, podemos encontrar semelhanças entre eles em determinados indicadores, e certas diferenças ao analisar outros. O cluster 1 por exemplo, possui liquidez saudável, retorno sobre capital investido bom, alto valor patrimonial e preço/lucro alto, enquanto o grupo 2 também apresentou liquidez saudável, alto retorno sobre o capital investido, preço/lucro negativo, o que indica prejuízo da empresa e valor patrimonial negativo. Desta maneira, é plausível dizer que o perfil de investidor para o grupo 1 é conservador, enquanto o perfil de investidor para o grupo de empresas 2 é moderado, pois possui algumas características de um perfil conservador, com alguns itens mais arriscados.

Agora, para o grupo 4 com 3 clusters, no qual encontramos 0,87 no Silhouette Score:

QTD_Cluster	N_cluster	Indicador	Valor
3	1	CAGR Receita 5 anos	13.06
3	2	CAGR Receita 5 anos	0.010000
3	3	CAGR Receita 5 anos	3.75
3	1	Margem líquida	120.75
3	2	Margem líquida	0.00
3	3	Margem líquida	639.59
3	1	VPA	25.18
3	2	VPA	0.000000
3	3	VPA	-82.95
3	1	P/ATIVOS	1.21
3	2	P/ATIVOS	33255.41

3	3	P/ATIVOS	0.172
3	1	dívida líquida / EBIT	2.12
3	2	dívida líquida / EBIT	0.000000
3	3	dívida líquida / EBIT	59.895

Tabela 5: Resultados média de valores para grupo 4 com 3 clusters
Fonte: Autores

Ao analisar os valores da tabela do grupo 4 dividido em 3 clusters, podemos rapidamente inferir que o grupo 2 trata-se de empresas para perfis mais arrojados, uma vez que possui 0,01 de CAGR de receita, ou seja, é uma empresa nova sem valores de receita mais velhos que 5 anos, possui baixíssimo valor patrimonial, baixa margem líquida e altíssimo valor de preço/ativos. Pelos indicadores desfavoráveis, para um investidor investir neste grupo de empresas, deve conter algum tipo de informação externa, como conhecimento do setor ou de seus fundadores. Investimentos com esse tipo de embasamento são comuns à investidores arrojados. Em seguida, ao avaliar o grupo 1, é possível ver um valor saudável de CAGR de receita, altíssima margem líquida, alto valor patrimonial, valor de preço/ativos saudável e dívida líquida / ebit em patamar saudável. O grupo 3, por sua vez, possui valor de CAGR saudável, porém menor que o grupo 1, indicando uma empresa com crescimento menor nos últimos 5 anos, um valor patrimonial negativo e valores de dívida líquida/ebit bem altos, indicando que o grupos 3 de ações conta com bons valores, entretando ainda alguns riscos ao se considerar determinados indicadores, assim, é plausível atribuir o grupo 3 de ações a um perfil de investidor moderado, e o perfil 1 de ações a um perfil de investidor mais conservador.

Para finalizar a discussão, trataremos uma análise da diferença dos indicadores de 4 ações públicas e 4 ações privadas, comparando a mesma quantidade de ações de cada setor. Público (Petrobras Ordinária: PETR3, Petrobras Preferencial: PETR4, Banco do Brasil Ordinária: BAS3, Banco Bradesco Ordinária: BBDC3, Banco Itaú Ordinária: ITUB3). Portanto, estão sendo consideradas duas petrolíferas públicas, 2 privados, 2 bancos públicos e 2 privados, com os seguintes dados:

TICKER	P/L	MARG. LIQUIDA	ROIC	LIQUIDEZ MEDIA DIARIA
PETR3	2,46	28,48	30,84	437.509.807,48
PETR4	2,21	28,48	30,84	1.746.547.284,17
BBAS3	4,4	12,67	-	592.634.410,38
BRSR3	6,24	6,59	-	80.820,07

PRI03	8,36	46,02	18,88	463.318.130,69
RPMG3	-0,15	-19,13	24,39	54.852,10
BBDC3	7,93	9,1	-	98.615.658,76
ITUB3	7,57	10,35	-	12.247.009,86

Tabela 6: Resultados de valores para 2 grupos de 4 ações cada.
Fonte: Autores

Através de uma análise da média dos grupos, público e privado, foi possível associar os valores dos indicadores referentes ao grupo 2: liquidez média diária, Preço / Lucro, margem líquida, ROIC, VPA.

E seus respectivos valores de média:

TICKER	P/L	MARG. LIQUIDA	ROIC	LIQUIDEZ MEDIA DIARIA
Média Público	3,8275	19,055	30,84	694.193.080,53
Média Privado	5,9275	11,585	21,635	143.558.912,85

Tabela 7: Resultados da média dos valores para 2 grupos de 4 ações cada.
Fonte: Autores

É possível ver que as empresas públicas possuem maior liquidez, maior retorno sobre capital investido, maior margem líquida e menor valor de preço/lucro, todos indicadores de perfil de investidor mais conservador, como explicado anteriormente neste artigo. Em contrapartida, as empresas privadas possuem menor margem líquida, menor liquidez diária, menor retorno sobre capital investido e maior valor de preço/lucro, todos valores que indicam empresas para pessoas que assumem mais riscos. Neste teste, os valores não são tão distintos quanto os grupos que foram avaliados, pois neste teste todas as empresas que estão sendo consideradas são de grande porte, e já possuem mercados consolidados, mesmo assim, é possível verificar que em uma análise comparativa, as lógicas se aplicam, e o setor público, que é sabidamente mais seguro por possuir a retaguarda pública, são mais coerentes com o perfil de investidor conservador.

VIII. CONCLUSÃO

Ao avaliar os resultados obtidos, portanto, é possível concluir que o algoritmo realiza o agrupamento e a clusterização das ações da base de dados de acordo com os indicadores considerados, e estes clusters podem ser coerentemente atribuídos aos perfis de investidor apresentados anteriormente neste artigo. Vale ressaltar que para 2 clusters, o algoritmo é mais eficiente em sua separação, como mostra o valor do silhouette score, e como é possível perceber nos valores das médias dos indicadores apresentados, enquanto com 3 clusters, ele também é eficiente, mas com maior proximidade nos valores das médias entre os grupos atribuídos a investidores conservadores e moderados.

Por último, é imprescindível destacar que o grupo de ações retornado pelo algoritmo não se trata de uma recomendação de compra, pois tal atitude só é permitida para profissionais com a certificação CNPI, o algoritmo recomenda uma indicação de estudos por parte dos investidores, para que estes possam realizar suas análises de maneira direcionada. Além disso, também destacamos que toda análise de ativos no mercado financeiro deve ser feita contemplando mais de um indicador, além de análises macroeconômicas, bem como objetivos pessoais.

REFERÊNCIAS

As referências utilizadas neste estudo serviram como bases para a formulação da tese e da motivação propriamente dita deste projeto. Serão enumeradas, a seguir, as referências utilizadas tanto na contextualização teórica, como na introdução e na apresentação de trabalhos existentes com temas semelhantes. Vale ressaltar que os 10 artigos citados anteriormente no item III deste relatório também atuam como referências neste projeto.

- [1] B3, BORA INVESTIR. (2023). Argentina eleva juros para 97% ao ano para conter inflação [Online]. Disponível em: <https://borainvestir.b3.com.br/noticias/mercado/argentina-eleva-os-juros-para-97-ao-ano-para-conter-a-inflacao/>
- [2] STUMPF, K., Curso preparatório para certificação CEA. TOPINVEST, 2022
- [3] STUMPF, K. Curso preparatório para certificação ANCORD. TOPINVEST, 2022.