

João Victor Rodrigues Ramalho

Aplicabilidade de técnicas de redução de dimensionalidade no índice Bovespa (Ibovespa B3)

Campina Grande - PB

2 de setembro de 2022

João Victor Rodrigues Ramalho

Aplicabilidade de técnicas de redução de dimensionalidade no índice Bovespa (Ibovespa B3)

Projeto de Trabalho de Conclusão de Curso
para obtenção do título de bacharel em En-
genharia Elétrica pela Universidade Federal
de Campina Grande.

Universidade Federal de Campina Grande - UFCG
Centro de Engenharia Elétrica e Informática - CEEI
Departamento de Engenharia Elétrica - DEE

Campina Grande - PB
2 de setembro de 2022

AGRADECIMENTOS

Agradeço em primeira instância a Deus por me dar força, saúde e resiliência, a minha família em especial a minha mãe Suênia, meu pai Nailson, minha irmã Juliana, minha tia Noalda e meu tio Jackson e meus dois primos Maria Julia e João Lucas, minha tia e madrinha Sueide que acompanharam de perto toda minha trajetória acadêmica, a minha segunda família a “Batfamília” e meus amigos Tiago, Wendel, Wesley, Caio, André, Leticia Dala e Brenda Seixas.

Agradeço também a toda família PET-Elétrica, grupo no qual foi essencial para meu desenvolvimento e crescimento, em especial aos amigos que a graduação me proporcionou, Pedro Jacob, Matheus Rique , Pedro Henrique e Francisco.

Agradeço os meus colegas de trabalho Rafael Poppe, Fausto Martins, Gabriel Moreira, e demais integrantes da Kapitalo Investimentos.

Agradeço a minha namorada e companheira Maria Alice, que acompanhou de perto todas as noites mal dormidas para o desenvolvimento deste trabalho.

Agradeço ao meu orientador Edmar Candeia Gurjão e todo corpo docente do Departamento de Engenharia Elétrica (DEE- UFCG).

RESUMO

A dimensionalidade é conhecida como o número de variáveis de entrada ou características presentes em um espaço de dados, e o processo de tratar a quantidade de características representantes desse espaço é chamado redução de dimensionalidade. A redução da dimensionalidade tem como objetivo converter um espaço de dados de uma determinada dimensão em uma outra inferior, de forma a assegurar que ainda forneça informação equivalente. Estas técnicas são amplamente utilizadas na modelagem do aprendizado de máquina para obter um modelo preditivo adequado, resolvendo ao mesmo tempo os problemas de classificação e regressão. Neste trabalho, foi realizado a implementação de algumas técnicas de redução de dimensionalidade, no principal índice de ações da bolsa de valores do Brasil, o índice Bovespa (Ibovespa B3), com auxílio da linguagem de programação Python.

Palavras-chaves: Álgebra Linear, Bovespa, B3, PCA, SVD, Python.

ABSTRACT

Dimensionality is known as the number of input variables or features present in a data space, and the process of addressing the number of features representing that space is called dimensionality reduction. Dimensionality reduction aims to convert a data space of a certain dimension into one that is lower in order to ensure that it still provides equivalent information. These techniques are widely used in machine learning modeling to obtain a suitable predictive model while solving classification and regression problems. In this work, the implementation of some dimensionality reduction techniques, on the main stock index of the Brazilian stock exchange, the Bovespa index (Ibovespa B3), was carried out with the aid of the Python programming language.

Keywords: Linear Algebra, Bovespa, B3, PCA, SVD, Python.

LISTA DE ILUSTRAÇÕES

Figura 1 – Relação dos códigos das ações e seus tipos	12
Figura 2 – Ilustração da composição da carteira do Ibovespa com foco nas 10 principais ações	13
Figura 3 – Representação visual da primeira componente principal	19
Figura 4 – Visualização da decomposição em valores singulares	21
Figura 6 – Ranking de compressão de uma imagem por meio do SVD	22
Figura 5 – Aproximação de truncamento para o SVD	22
Figura 7 – Representação dos diferentes valores nominais das ações	24
Figura 8 – Matriz de correlação das ações do índice Ibovespa	26
Figura 9 – Variância explicada das 10 principais componentes	28
Figura 10 – Matriz de correlação ajustada em 3 dimensões	29
Figura 11 – Comparativo entre o retorno do iBovespa versus o Ibovespa Ponderado	31
Figura 12 – Gráfico pizza da Carteira Teórica Obtida pelo PCA - Metodo 1	34
Figura 14 – Gráfico pizza da Carteira Teórica Obtida pelo PCA - Metodo 2	36
Figura 15 – Comparação do retorno da carteira do Ibovespa e o portfolio obtido com a aplicação do PCA - método 2	37
Figura 16 – Ilustração da carteira reduzida com o método de Decomposição em Valores Singulares em gráfico pizza	39
Figura 13 – Comparação do retorno da carteira do Ibovespa e o portfolio obtido com a aplicação do PCA - método 1	40
Figura 17 – Retorno observado do SVD Portfólio ao lado do retorno do Ibovespa Ponderado	41
Figura 18 – Matriz de correlação: Bens Industriais	45
Figura 19 – Matriz de correlação: Bens Industriais	46
Figura 20 – Matriz de correlação: Comunicações	47
Figura 21 – Matriz de correlação: Consumo Cíclico	48
Figura 22 – Matriz de correlação: Consumo Cíclico	49
Figura 23 – Matriz de correlação: Financeiro	50
Figura 24 – Matriz de correlação: Petróleo, Gás e Biocombustíveis	51
Figura 25 – Matriz de correlação: Saúde	52
Figura 26 – Matriz de correlação: Tecnologia da Informação	53

LISTA DE TABELAS

Tabela 1 – Representação a matriz normalizada de retornos	24
Tabela 2 – Composição da Carteira Teórica Obtida pelo Metodo 1	33
Tabela 3 – Composição da Carteira Teórica Obtida pelo PCA - Metodo 2	35
Tabela 4 – Carteira Reduzida em 28 ações com o método de Decomposição em Valores Singulares	38

LISTA DE ABREVIATURAS E SIGLAS

PCA *Principal Component Analysis*

SVD *Singular Value Decomposition*

B3 Brasil, Bolsa, Balcão

LISTA DE SÍMBOLOS

Λ	Letra grega Lambda
λ	Letra grega lambda minúscula
Σ	Letra grega Sigma
σ	Letra grega sigma minúscula

SUMÁRIO

1	INTRODUÇÃO	11
1.1	O índice Bovespa	11
1.2	Objetivos	13
1.2.0.1	Objetivo Geral	13
1.2.0.2	Objetivo Específico	13
1.3	Motivação	14
2	CONCEITOS BASE	15
2.1	Autovetores e Autovalores	15
2.2	Variância	15
2.3	Covariância	16
3	REDUÇÃO DE DIMENSIONALIDADE	17
3.1	Análise de Componentes Principais	17
3.1.1	Método	17
3.2	DECOMPOSIÇÃO EM VALORES SINGULARES	20
3.2.1	Método	20
4	APLICAÇÃO NO IBOVESPA	23
4.1	Configurações iniciais	23
4.2	Tratamento de dados	23
4.3	Redução em Componentes Principais - Abordagem 1	27
4.4	Redução em Componentes Principais - Abordagem 2	28
4.5	DECOMPOSIÇÃO EM VALORES SINGULARES	30
4.6	Resultados	31
4.6.1	Considerações Essenciais	31
4.6.2	Método 1	32
4.6.3	Método 2	34
4.7	Decomposição em valores Singulares	37
5	CONCLUSÃO	42
	REFERÊNCIAS	43

ANEXOS	44
ANEXO A – CORRELAÇÃO SETORIAL	45

1 INTRODUÇÃO

O termo dimensionalidade representa um agrupamento de característica que representa um padrão. Ao se mapear o espaço de características, é possível implementar algoritmos que são capazes reduzir as dimensões da representação deste espaço, sem que ele perca sua essência. Equações com um elevado número de variáveis de entrada (características) são objetos perfeitos a serem aplicados às técnicas de redução de dimensionalidade, devido ao fato de que algumas dessas variáveis podem ser redundantes.

O índice Ibovespa é um indicador que mede o desempenho geral das principais ações da Bolsa de Valores do Brasil (B3), o índice é calculado com base na performance das ações mais negociadas na bolsa, sendo composto atualmente por 90 ações, sofrendo alterações a cada 4 meses (rebalanceamento), com objetivo de representar o cenário atual das empresas de capital aberto. Devido à sua característica de tentar representar as ações mais negociadas da bolsa de valores, acaba possuindo um aspecto redundante em sua representatividade, possuindo várias empresas do mesmo setor que possuem comportamentos semelhantes no decorrer da história

1.1 O ÍNDICE BOVESPA

Um índice de mercado é definido como a evolução de uma carteira de ações pré-determinada por uma certa metodologia, em analogia é possível comparar com um índice inflacionário, como o IGPM (Índice Geral de Preços de Mercado), no qual é composto por uma cesta de produtos, sendo ponderado pelo comportamento dos preços dos mesmos. A diferença, contudo, é que os índices inflacionários observam a evolução média de uma cesta de produtos e/ou serviços, enquanto os índices do mercado de ações, por outra ótica, observam a performance média ponderada de ações da bolsa de valores.

Para entendermos o comportamento dos índices é necessário primeiramente compreender o que são as ações, o termo representa frações de participações em companhias de capital aberto. As ações ordinárias (ON) garantem que o seu possuidor tenha o direito ao voto nas assembleias da empresa. As preferenciais (PN) não têm direito a voto, mas entrega uma certa prioridade no momento de receber os proventos que possam ser distribuídos. Na bolsa de valores as companhias são representadas por um *ticker*, abreviaturas do nome da companhia, adicionado de um código no final, como por exemplo PETR4, é o ticker que representa a Petrobrás e o código 4, representa que se trata de uma ação preferencial, a relação dos códigos e os seus significados pode ser observado na [Figura 1](#)

As bolsas de valores pelo mundo possuem os seus principais índices representativos



Código	Tipo
1	Direito de Subscrição - Ação Ordinária
2	Direito de Subscrição - Ação Preferencial
7	Preferenciais Classe C
8	Preferenciais Classe D
9	Recibo de Subscrição - Ação Ordinária
10	Recibo de Subscrição - Ação Preferencial

Figura 1 – Relação dos códigos das ações e seus tipos

Fonte: (VIEIRA, 2017)

e suas determinadas metodologias, no caso dos Estados Unidos, as bolsas são representadas pelos índices SP 500, *Dow Jones Industrial*, *NASDAQ Composite*, já a maior bolsa de valores fora dos Estados Unidos, é a bolsa de Tokyo, representada pelo índice Nikkei 225. A bolsa de valores brasileira (B3) antiga BMFBovespa, é a maior bolsa da América Latina e a única bolsa brasileira em operação e a sua capitalização de mercado passa de 4 trilhões de reais, e é representada pelo índice Bovespa (Ibovespa B3)

A metodologia atual do Ibovespa foi elaborada em conjunto com vários participantes e segmentos do mercado visando mantê-lo como o índice que mais representa com exatidão o desempenho do mercado brasileiro. Como composição atual, o índice é formado por 90 ações, entretanto, 10 ações conseguem representar mais de 50%, como pode ser observado na [Figura 2](#)

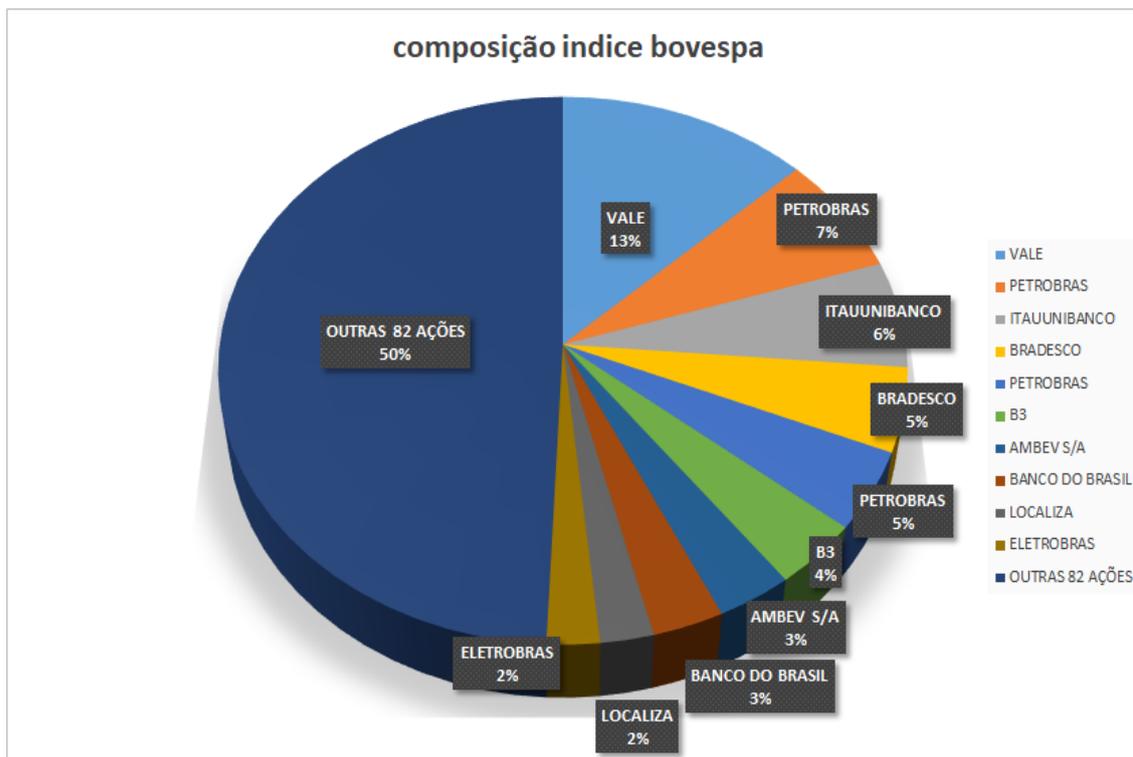


Figura 2 – Ilustração da composição da carteira do Ibovespa com foco nas 10 principais ações

Fonte: Próprio Autor.

1.2 OBJETIVOS

1.2.0.1 OBJETIVO GERAL

O presente trabalho tem como objetivo aplicar técnicas de redução de dimensionalidade nas ações presentes no índice iBovespa

1.2.0.2 OBJETIVO ESPECÍFICO

De maneira específica, o presente trabalho busca realizar uma replicação no comportamento do índice Ibovespa, de forma que possa ser utilizado em aplicações específicas de proteção de investimentos (*Hedge*), sendo abstraído o retorno nominal da carteira teórica idealizada, e não sendo uma recomendação de investimento. A metodologia para custear operações na bolsa brasileira torna inviável a sua utilização a fins de arbitragem, uma vez que os custos envolvidos nas operações no Brasil possuem influência do preço e do volume de ativos operados.

1.3 MOTIVAÇÃO

Analisando a carteira atual do iBovespa, é perceptível que a sua metodologia traz uma certa redundância de representatividade, tal que existem empresas de setores semelhantes com comportamentos históricos semelhantes e com pesos distintos. Esse trabalho tenta tornar possível a recriação de uma carteira na qual consiga representar o mercado de ações brasileiro com menos ativos e reduzindo o risco de concentração, aplicando conhecimentos técnicos adquiridos em disciplinas como Álgebra Linear, Processos Estocásticos, Economia, Controle Analógico, Controle Digital e Informática Industrial obtidos na graduação de Engenharia Elétrica.

2 CONCEITOS BASE

Neste capítulo serão apresentados os principais conceitos e processos para a aplicação de técnicas de redução de dimensionalidade

2.1 AUTOVETORES E AUTOVALORES

Os autovalores de uma matriz também são chamados de valor próprio ou valor característico. Com objetivos de entender sua definição, é analisado uma matriz A quadrada (Equação 2.1)

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \quad (2.1)$$

Multiplicando a matriz A por um vetor $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$ não nulo, obtém-se um outro

vetor também de dimensão $m \times 1$. Por outro lado, ao multiplicar o mesmo vetor \mathbf{v} por uma constante λ , também é obtido como resultado um vetor de dimensão $m \times 1$, sendo assim, caso a relação seja satisfeita: $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$ É possível definir λ como um **Autovalor** da matriz A , e que \mathbf{v} é um **Autovetor** associado ao escalar λ . Os autovetores são chamados também por "vetores próprios" ou "vetores característicos" (BOLDRINI et al., 1980).

2.2 VARIÂNCIA

A Variância (Equação 2.2) de uma variável aleatória é uma medida dispersão estatística, indicando quão distante em geral os seus valores se encontram do valor esperado (MONTGOMERY; RUNGER; CALADO, 2000)

$$\text{Var}(X) = E[(X - \mu)^2] \quad (2.2)$$

Sendo $E[X]$ o valor esperado, definido pela Equação 2.3

$$E[X] = \sum_{i=1}^{\infty} x_i p_i \quad (2.3)$$

Onde x_1, x_2, \dots são os resultados possíveis da variável aleatória X e p_1, p_2, \dots as probabilidades correspondentes.

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.4)$$

O desvio padrão (Equação 2.4) é uma medida que expressa o grau de dispersão de um conjunto de dados, ou seja, o desvio padrão indica a uniformidade de um determinado conjunto de dados que é definido como a raiz quadrada da variância, sendo uma medida unidimensional e tem como variáveis os itens abaixo:

- s : Desvio padrão
- x_i : Valor analisado
- \bar{x} : Média Aritmética do conjunto
- N : Número de dados do conjunto

2.3 COVARIÂNCIA

A covariância (Equação 2.5) é uma medida estatística em que é possível comparar duas variáveis, permitindo entender como elas se relacionam entre si. Tem característica bidimensional e é calculada de maneira independente, em que, a covariância de variáveis independentes será zero (MONTGOMERY; RUNGER; CALADO, 2000).

$$K_{X_i X_j} = \text{cov}[X_i, X_j] = E[(X_i - E[X_i])(X_j - E[X_j])] \quad (2.5)$$

Quando se trata de um conjunto de dados, realiza-se um comparativo com todo o conjunto, para isso, é necessário aplicarmos uma matriz de covariância (Equação 2.6), sendo aplicável em sua diagonal a propriedade visualizada pela Equação 2.7, no qual o comparativo entre duas variáveis idênticas se define como sua própria dispersão.

$$C = \begin{bmatrix} \text{cov}(X, X) & \text{cov}(X, Y) & \text{cov}(X, Z) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) & \text{cov}(Y, Z) \\ \text{cov}(Z, X) & \text{cov}(Z, Y) & \text{cov}(Z, Z) \end{bmatrix} \quad (2.6)$$

$$\text{cov}(\mathbf{X}, \mathbf{X}) = \text{var}(\mathbf{X}) = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T]. \quad (2.7)$$

3 REDUÇÃO DE DIMENSIONALIDADE

Neste capítulo serão apresentadas as técnicas de redução de dimensionalidade utilizadas para a validação de sua aplicabilidade no Índice Bovespa, sendo elas: Redução em componentes principais e a decomposição em valores singulares.

3.1 ANÁLISE DE COMPONENTES PRINCIPAIS

A técnica conhecida como Análise de Componentes Principais, advinda do inglês "*Principal Component Analysis*" (PCA), é definida como um dos maiores resultados da álgebra linear (SHLENS, 2014), sendo utilizada de forma ampla na neurociência e na computação, devido ao fato de ser um método simples e não paramétrico de extração de dados relevantes em um conjunto de dados. Portanto, é aplicado uma redução do conjunto que busca revelar um subconjunto com relevância significativa, conseguindo assim, representar o conjunto por completo. Os conceitos teóricos e o método de aplicação serão abordados nessa seção.

3.1.1 MÉTODO

Tomando como base a matriz de dados genérica apresentada na [Equação 2.1](#), de m características (colunas) e n indivíduos (linhas), a estrutura de interdependência das variáveis é apresentada pela matriz de covariância. A partir da matriz \mathbf{A} de dados de ordem 'm x n', é possível fazer uma estimativa da matriz de covariância que representaremos por \mathbf{S} . A matriz \mathbf{S} é simétrica e de ordem 'm x m'.

$$S = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots \text{Cov}(X_1, X_3) & \cdots \text{Cov}(X_1, X_m) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots \text{Cov}(X_2, X_3) & \cdots \text{Cov}(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \text{Cov}(X_n, X_3) & \cdots \text{Var}(X_m) \end{bmatrix} \quad (3.1)$$

Corriqueiramente os dados podem apresentar unidades de medidas diferentes, e por isso, é necessário realizar um processo de padronização nos dados (VARELLA, 2008), na forma prospota na [Equação 3.2](#)

$$Z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s(x_j)}. \quad (3.2)$$

em que $i = 1, 2, 3 \dots n$, $j = 1, 2, 3 \dots m$ e $s(x_j) = \sqrt{\text{var}(X_j)}$ e \bar{x}_j a média das amostras.

Depois de aplicada a padronização, é possível obter uma nova matriz de dados (Equação 3.3).

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \quad (3.3)$$

A matriz \mathbf{Z} das variáveis padronizadas z_j é semelhante a matriz de correlação na Equação 3.1.1. Para determinar as componentes principais, comumente partimos da matriz de correlação \mathbf{R} . É importante salientar que o resultado final que será encontrado para a análise a partir da matriz \mathbf{S} pode ser diferente do resultado encontrado a partir da matriz \mathbf{R} . Recomenda-se que a normalização deva ser feita apenas quando as unidades de medidas das características observadas não forem compatíveis.

$$\mathbf{R} = \begin{bmatrix} 1 & \frac{E[(X_1-\mu_1)(X_2-\mu_2)]}{\sigma(X_1)\sigma(X_2)} & \cdots & \frac{E[(X_1-\mu_1)(X_n-\mu_n)]}{\sigma(X_1)\sigma(X_n)} \\ \frac{E[(X_2-\mu_2)(X_1-\mu_1)]}{\sigma(X_2)\sigma(X_1)} & 1 & \cdots & \frac{E[(X_2-\mu_2)(X_n-\mu_n)]}{\sigma(X_2)\sigma(X_n)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{E[(X_n-\mu_n)(X_1-\mu_1)]}{\sigma(X_n)\sigma(X_1)} & \frac{E[(X_n-\mu_n)(X_2-\mu_2)]}{\sigma(X_n)\sigma(X_2)} & \cdots & 1 \end{bmatrix} \quad (3.4)$$

Caso a matriz \mathbf{R} não apresente nenhuma coluna que seja combinação linear de outra, tornando satisfeita a equação $|\mathbf{R} - \lambda I| = 0$, que por sua vez, terá ‘p’ raízes chamadas de autovalores ou raízes características. No processo de montagem da matriz de dados é importante observar que o valor de ‘n’ (características) deve ser pelo menos igual a ‘p+1’, uma vez que, é desejável construir um experimento para analisar o comportamento de ‘p’ características. Recomenda-se que o matriz de referência apresente no mínimo, ‘p+1’ tratamentos (VARELLA, 2008)

Sendo $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m$ as raízes da equação característica da matriz \mathbf{R} ou \mathbf{X} , então

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_m \quad (3.5)$$

Para cada autovalor λ_i existe um autovetor \mathbf{v} , como foi demonstrado na seção 2.1

Sendo \mathbf{v}_i o autovetor correspondente ao autovalor λ_i , então o i-ésimo componente principal é dado na Equação 3.6

$$\mathbf{Y}_i = a_{i1}\mathbf{v}_1 + a_{i2}\mathbf{v}_2 + a_{i3}\mathbf{v}_3 + \dots + a_{in}\mathbf{v}_m. \quad (3.6)$$

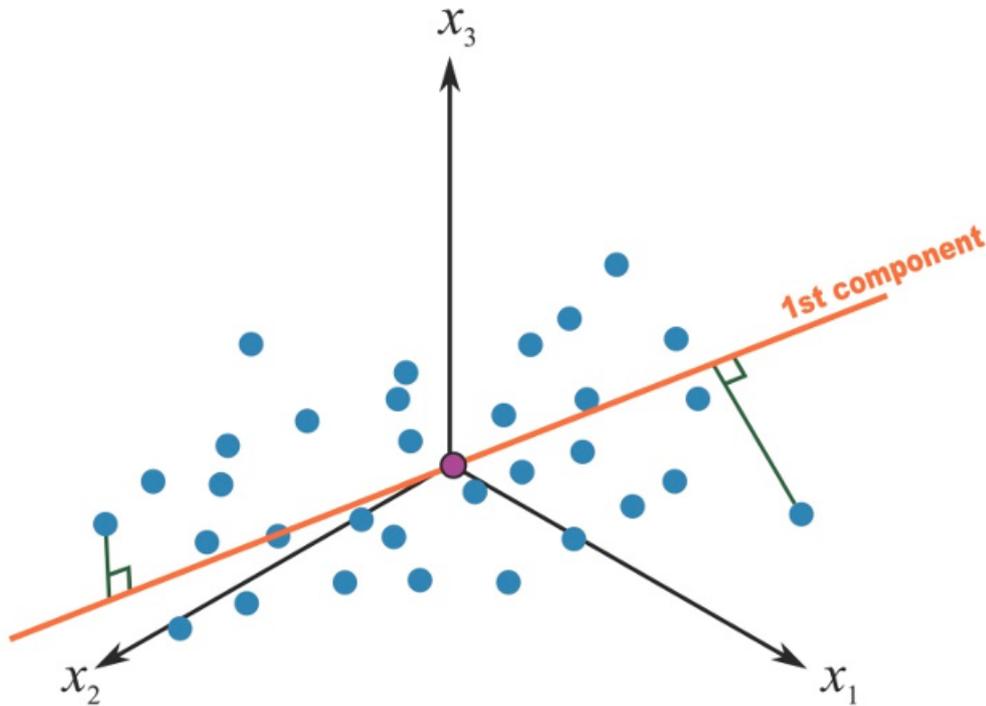


Figura 3 – Representação visual da primeira componente principal

Fonte: (DUNN, 2019)

É possível visualizar uma representação do que seria a primeira componente principal de um conjunto genérico pela Figura 3. As componentes principais usualmente são denominadas pela sigla "PC*i*", no qual *i* representa o numeral da componente a ser exposta, sendo PC0 e PC1 a primeira e segunda componente respectivamente.

- $\text{Var}(\mathbf{Y}_i) = \lambda_i$
- $\text{Var}(\mathbf{Y}_1) > \text{Var}(\mathbf{Y}_2) > \text{Var}(\mathbf{Y}_3) > \dots > \text{Var}(\mathbf{Y}_m)$
- O total de variância é igual ao somatório dos autovalores que é igual ao total de variância dos componentes principais
- Os componentes principais não são correlacionados entre si: $\text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = 0$

A contribuição \mathbf{B}_i de cada componente principal \mathbf{Y}_i é expressa em porcentagem. Calcula-se dividindo a variância de \mathbf{Y}_i pela variância total (ALDRIDGE; AVELLANEDA, 2021).

$$\mathbf{B}_i = 100 * \frac{\text{Var}(\mathbf{Y}_i)}{\sum_{i=1}^p \text{Var}(\mathbf{Y}_i)} \quad (3.7)$$

A importância de uma componente principal é avaliada pelo seu grau de relevância na matriz de dados original, isso pode ser observado pelo percentual de variância explicada acumulada pelo componente. A soma dos primeiros n Autovalores representa a proporção de informação retida na redução de dimensão.

Partindo dessa análise, é possível decidir o número de componentes que vamos utilizar, isto é, quantos componentes serão utilizados para diferenciar os indivíduos. Até o presente momento, não existem métodos estatísticos que possam auxiliar na escolha da quantidade de componentes.

3.2 DECOMPOSIÇÃO EM VALORES SINGULARES

A decomposição em valores singulares, do inglês *Singular Values Decomposition* (SVD), é uma das ferramentas mais poderosas da álgebra linear para o tratamento de dados, utilizada largamente para compressão de imagens de alta resolução, reduzindo a quantidade de memória alocada para o seu armazenamento. É considerada o primeiro passo para outras técnicas de redução de dimensionalidade, podendo ser considerada como uma generalização da Transformada de Fourier (RUDY et al., 2019). Nesta seção, será tratado a intuição matemática por trás do método que embasa o SVD.

3.2.1 MÉTODO

O método de decomposição em valores singulares é fundamentado de forma que uma matriz de dados \mathbf{A}_{mm} , possa ser fatorada de acordo com Equação 3.8, sendo \mathbf{U}_{mm} e Σ matrizes de rotação. Sendo essas condições satisfeitas, é possível realizar a decomposição da matriz original em valores singulares

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (3.8)$$

O lado direito da Equação 3.8 mostra a sua decomposição em valores singulares, analisando o conjunto de características (colunas) da matriz \mathbf{A} , sendo σ_i o valor singular da posição i^{th} , \mathbf{u}_i e \mathbf{v}_i as colunas de \mathbf{U} e \mathbf{V} , na posição i^{th} (Equação 3.9) o aspecto somatório acrescenta uma característica de camadas, a partir dessa característica, é possível aplicar uma redução de dimensionalidade, por meio de um truncamento na dimensão original.

Entretanto, para que tudo isso seja considerado, é necessário que a matriz \mathbf{A} se encaixe nas condições de ser uma matriz quadrada e com autovalores reais

$$\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i \tag{3.9}$$

$$\mathbf{A}^*\mathbf{u}_i = \sigma_i\mathbf{v}_i \tag{3.10}$$

Para calcularmos o SVD, é necessário encontrar o conjunto necessário de valores e vetores singulares para construir \mathbf{U} e \mathbf{V} , Σ . Partindo do princípio que a [Equação 3.8](#) existe, temos as relações exibidas na [Equação 3.11](#)

$$\begin{aligned} \mathbf{A}^T\mathbf{A} &= \mathbf{V}\Sigma^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^* = \mathbf{V}(\Sigma^T\Sigma)\mathbf{V}^T \\ \mathbf{A}\mathbf{A}^T &= \mathbf{U}\Sigma\mathbf{V}^T\mathbf{V}\Sigma^T\mathbf{U}^T = \mathbf{U}(\Sigma\Sigma^T)\mathbf{U}^* \end{aligned} \tag{3.11}$$

Tendo como informações principais as seguintes relações:

- \mathbf{V} contém os autovetores autonormais a $\mathbf{A}^T\mathbf{A}$
- \mathbf{U} contém os autovetores autonormais a $\mathbf{A}\mathbf{A}^T$
- σ_i é a matriz diagonal, composta pelos valores singulares

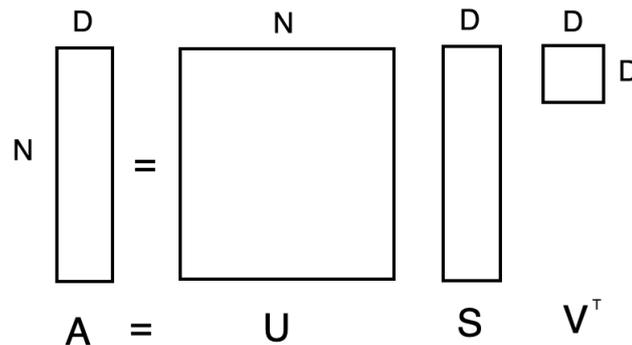


Figura 4 – Visualização da decomposição em valores singulares

Fonte: (RUDY et al., 2019)

De maneira visual, a [Equação 3.8](#), pode ser visualizada na [Figura 4](#), a redução de dimensionalidade ocorre de maneira a determinar a quantidade de valores singulares que será utilizado, diferentemente da técnica de análise de componentes principais, como pode ser observado, pela [Figura 5](#).

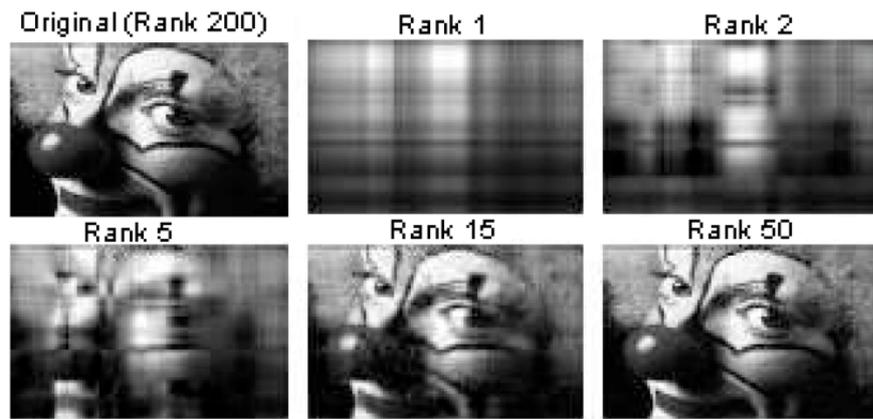


Figura 6 – Ranking de compressão de uma imagem por meio do SVD

Fonte: (RUDY et al., 2019)

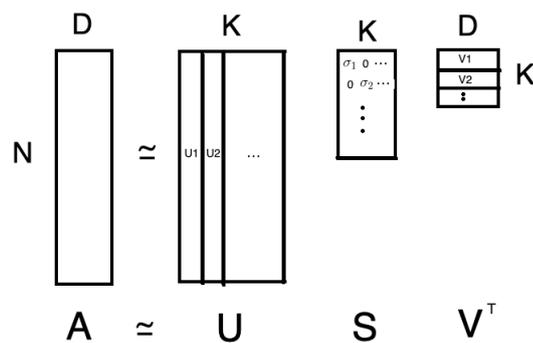


Figura 5 – Aproximação de truncamento para o SVD

Fonte: (RUDY et al., 2019)

Analisando de maneira simplista a [Figura 5](#), podemos definir a matriz \mathbf{U} como a matriz de Autovetores ordenados conforme a variância explicada, ou seja, o grau de relevância. Sendo \mathbf{S} , a matriz diagonal de valores singulares. O truncamento parte do pressuposto que a significância é relativa, e que pode ser escolhido apenas as k linhas que representariam a fidelidade da matriz original \mathbf{A} , considerando os valores subsequentes como ruído ([GEISS, 2006](#)). A partir disso é obtido um valor aproximado para a matriz \mathbf{A} . As consequências de escolha do truncamento podem ser visualizadas por meio da aplicação do SVD na compressão de uma imagem na [Figura 6](#)

4 APLICAÇÃO NO IBOVESPA

Nesta seção será apresentada a metodologia de aplicação da redução em componentes principais e decomposição em valores singulares no Índice Bovespa, por meio da linguagem Python.

4.1 CONFIGURAÇÕES INICIAS

O ambiente utilizado para a realização da implementação dos algoritmos foi o Jupyter Notebook, uma interface gráfica que permite a edição de *notebooks* em um navegador web, a linguagem de programação escolhida foi a linguagem Python, devido a quantidade de documentação registrada e os pacotes disponíveis. Como fonte de dados, foi utilizado o **B3 UP2DATA**, serviço de dados da própria bolsa de valores, contudo o estudo realizado pode ser replicado com fontes gratuitas de dados.

Os pacotes utilizados foram exibidos abaixo:

```
\%matplotlib inline
import pandas as pd
import numpy as np
from numpy.linalg import lstsq
from numpy.random import randn, seed
import matplotlib.pyplot as plt
from matplotlib.patches import Patch
from sklearn.decomposition import TruncatedSVD
from scipy import linalg
import seaborn as sns
from sklearn.decomposition import PCA
```

4.2 TRATAMENTO DE DADOS

O preço de uma ação tem um aspecto dinâmico e está completamente atrelado a uma série de fatores, dentre eles a quantidade de ações disponíveis em circulação (*Free Float*), sendo assim, é possível perceber que existem diferentes escalas de preços, como mostra a [Figura 7](#). Por isso, a fonte de dados principal será uma matriz de retornos logarítmicos, que representa a valorização ou desvalorização dos ativos em um determinado dia.

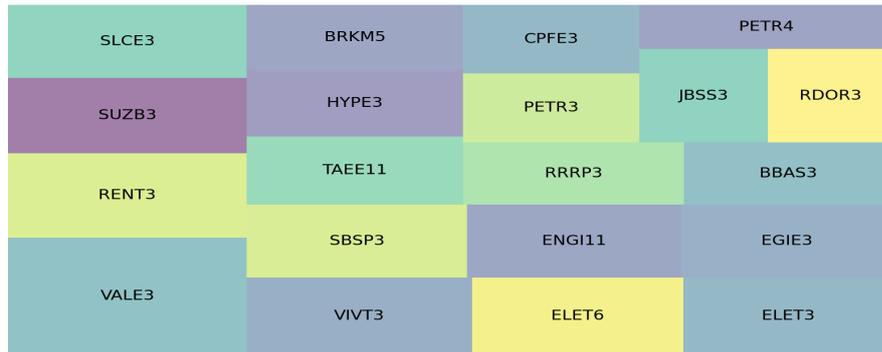


Figura 7 – Representação dos diferentes valores nominais das ações

Fonte: Próprio Autor.

Portanto, para obter a matriz de dados \mathbf{Z} , demonstrada na [Equação 3.1.1](#), foi utilizado o pacote *sklearn*, que aplica a [Equação 3.2](#) na matriz de retornos, obtendo assim uma matriz normalizada, exemplificada na [Tabela 1](#)

```

scaler = StandardScaler().fit(returns)
rescaledDataset = pd.DataFrame(scaler.fit_transform(returns), columns =
                                returns.columns, index = returns.
                                index)
returns.dropna(how='any', inplace=True)
rescaledDataset.dropna(how='any', inplace=True)
rescaledDataset[['ABEV3', 'ALPA4', 'PETR4', 'BBDC4']].head(4)

```

Ações	ABEV3	ALPA4	PETR4	BBDC4
Data				
2022-05-12	0.964926	-0.203819	0.935395	1.167477
2022-05-13	0.424884	0.375247	0.214835	0.392895
2022-05-16	0.162271	-0.363098	0.356288	0.765585
2022-05-17	0.580512	-0.118283	0.273434	1.125540

Tabela 1 – Representação a matriz normalizada de retornos

Fonte: Próprio Autor.

Por fim, antes de iniciar o processo de redução em componentes principais, é necessário encontrar a matriz de correlação, demonstrada na [Equação 2.1](#), para isso foi

utilizada uma função nativa do pacote **NumPy** (*numpy.corrcoef*), e se obteve a matriz final de correlação exibida em [Figura 8](#)

```
correlation = np.corrcoef(rescaledDataset)
```

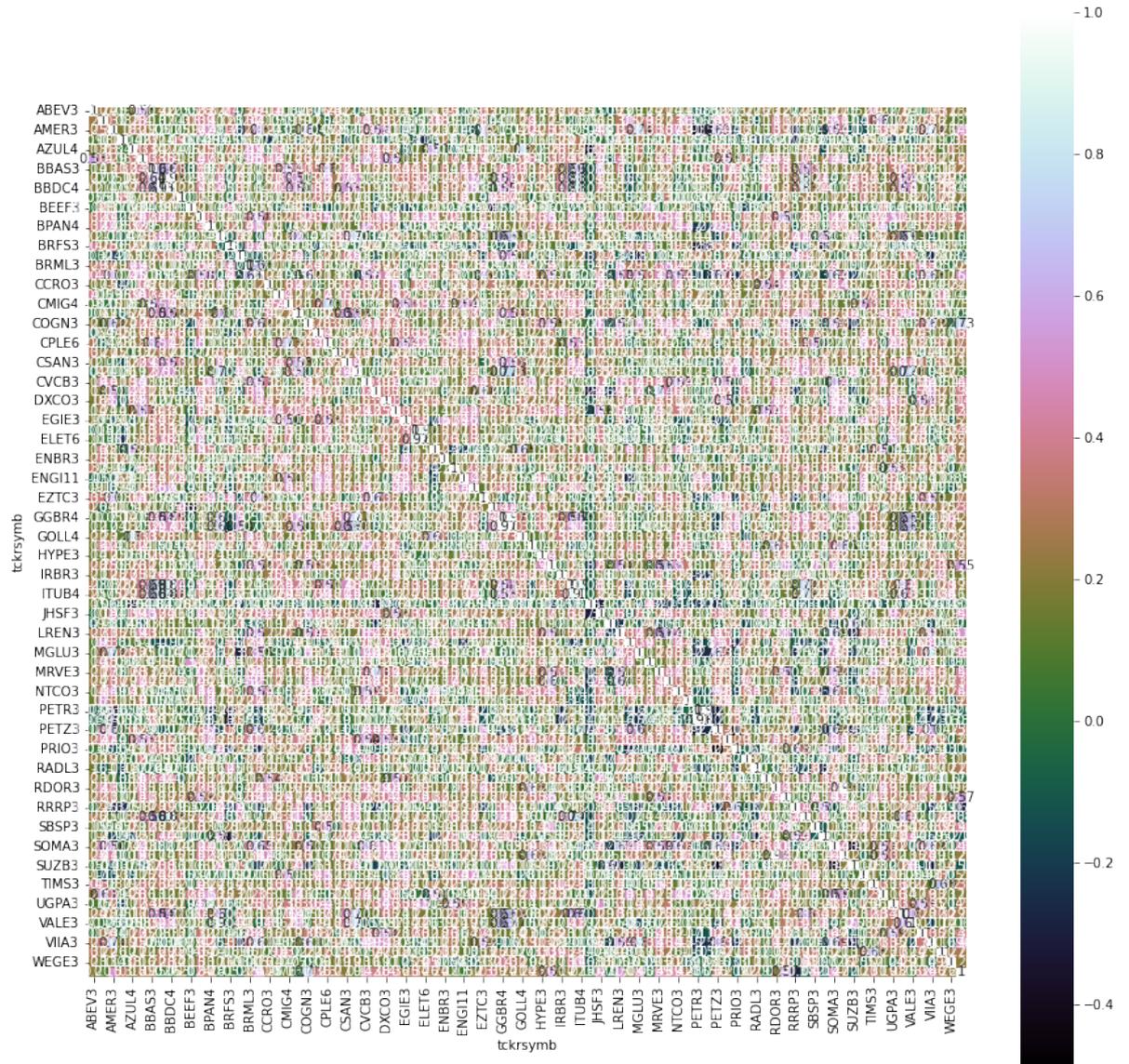


Figura 8 – Matriz de correlação das ações do índice Ibovespa

Fonte: Próprio Autor.

Devido ao fato do ibovespa possuir 90 ações, a visualização da sua matriz de correlação pode ser prejudicada. A caráter demonstrativo, o anexo A apresenta as matrizes

de correlações das ações que compõem o Ibovespa de forma setorizada, a título de visualizar o grau de correlação entre os ativos de mesmo setor.

4.3 REDUÇÃO EM COMPONENTES PRINCIPAIS - ABORDAGEM 1

Como primeira abordagem da redução em componentes principais, foi utilizado o pacote *sklearn*, em sua capacidade máxima de decomposição, utilizando como base matriz de correlação ajustada, sendo assim, como resultado foi retornada 65 componentes principais da matriz de correlação.

```
pca = PCA()
PrincipalComponent=pca.fit(correlation)
len(pca.components_)
# output 65
```

O método `pca.components_` é o conjunto de todos os Autovetores (um Autovetor para cada componente principal). Uma vez que você tenha os autovetores usando `pca.components_`, é possível obter os autovalores.

$$\mathbf{v} = \begin{bmatrix} -0.07 \\ -0.10 \\ -0.10 \\ -0.11 \\ -0.14 \\ \vdots \\ -0.08 \\ -0.12 \end{bmatrix} \quad (4.1)$$

Sendo \mathbf{v} (Equação 4.1) o autovetor da primeira componente principal, os seus autovalores podem ser assumidos como as variâncias do espaço reduzido (ALDRIDGE; AVELLANEDA, 2021)

Sendo a variância explicada o grau de significância das componentes (subseção 3.1.1), a proporção em que o modelo matemático se responsabiliza pela variação (dispersão) do espaço reduzido pode ser obtida pelo método `pca.explained_variance_` (4.2) (Figura 9)

$$[27.42 \quad 8.88 \quad 4.99 \quad 3.12 \quad 2.70 \quad 2.64 \quad 2.54 \quad \cdots 0.02 \quad 0.02 \quad 0.00] \quad (4.2)$$

Por fim, para obter a carteira reduzida, aplicou-se o código abaixo, no qual a variável n foi escolhida de forma empírica, sendo ajustada para 73 ativos para as condições específicas tratadas.

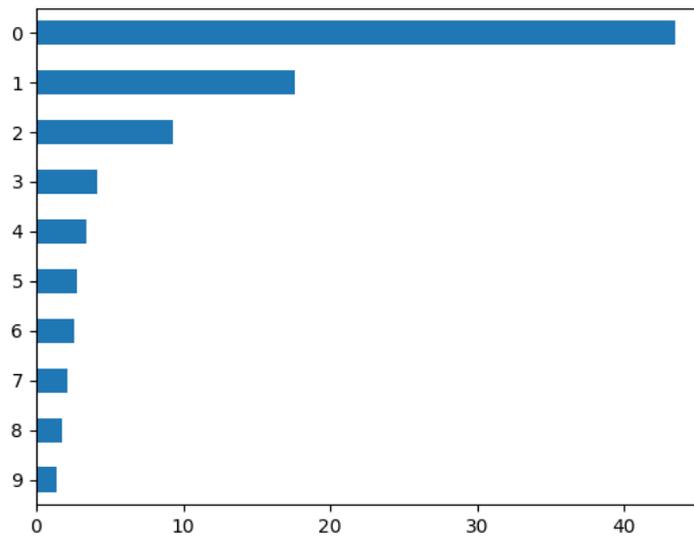


Figura 9 – Variância explicada das 10 principais componentes

Fonte: Próprio Autor.

```

componente_principal = pca.components_[0,:]

# Numero de Ativos
n = 73
bool_vector = abs(componente_principal) > np.sort(abs(
    componente_principal))[len(
    componente_principal)-n]
vetor_pesos = (componente_principal[bool_vector])
vetor_acoes = returns.columns[bool_vector]

```

Foi utilizada apenas a primeira componente principal, buscando minimizar a correlação entre alguns ativos do mercado (YANG, 2015)

4.4 REDUÇÃO EM COMPONENTES PRINCIPAIS - ABORDAGEM 2

Como segunda abordagem, foi forçado o algoritmo a reduzir toda a matriz de correlação ajustada, em apenas 10 componentes, diferentemente das 65 utilizadas na abordagem anterior. Na seção de resultados de considerações finais será explicada a principal diferença prática percebida ao variar o número da dimensão analisada.

Para a redução em apenas 10 componentes, é aplicado a atributo *n_components*

do pacote **PCA**

```
pca = PCA(n_components=10)
PrincipalComponent=pca.fit(X_train)
len(pca.components_)
# output 10
```

O objetivo dessa análise traz à tona a possibilidade de escolher o número de componentes, além de perceber o comportamento do modelo com um número inferior de componentes, resultando em uma variância explicada menor do que o ideal para a representação. A [Figura 10](#), mostra a interseção da correlação entre 3 componentes

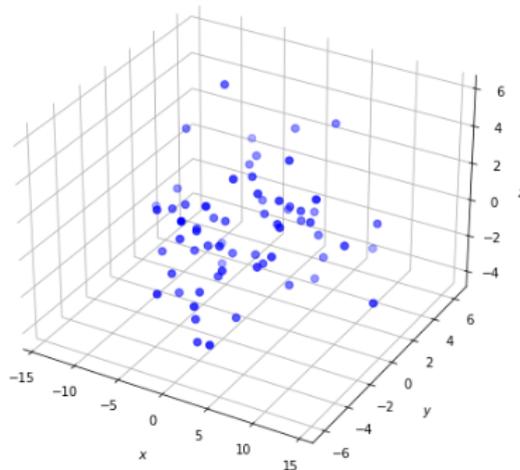


Figura 10 – Matriz de correlação ajustada em 3 dimensões

Fonte: Próprio Autor.

Para a análise dos Autovetores e Autovalores, de dimensão reduzida, pôde-se utilizar os mesmos métodos mencionados na primeira abordagem. Tendo como Autovetor da componente principal \mathbf{h} (Equação 4.3)

$$\mathbf{h} = \begin{bmatrix} -0.03 \\ -0.11 \\ -0.21 \\ -0.07 \\ \vdots \\ -0.01 \\ -0.04 \\ -0.17 \end{bmatrix} \quad (4.3)$$

Por fim, para obter a carteira reduzida, foi aplicado de forma empírica a variável n

```

componente_principal = pca2.components_[0,:]

# Numero de Ativos
n = 60
bool_vector = abs(componente_principal) > np.sort(abs(
    componente_principal))[len(
    componente_principal)-n]
vetor_pesos = (componente_principal[bool_vector])
vetor_acoes = returns.columns[bool_vector]

```

4.5 DECOMPOSIÇÃO EM VALORES SINGULARES

Para realizar a decomposição em valores singulares (SVD), aplica-se a função *svd* do pacote **NumPy**, se utiliza a [Equação 3.8](#) sobre a matriz de correlação, exibida na [Figura 8](#). O código utilizado para realizar a decomposição é exibido abaixo:

```

from numpy import linalg as LA

corr = np.corrcoef(returns1)
U, s, Vh = LA.svd(corr, full_matrices=False, hermitian = True)

```

Foi escolhido uma redução mais severa, trazendo o portfólio reduzido para 28 ações no total, apresentando um retorno de 68,89% do tamanho original da carteira do iBovespa. O método de seleção das ações foi obtido por meio do primeiro Autovetor da matriz de correlação ([ALDRIDGE; AVELLANEDA, 2021](#)), já para o cálculo dos pesos, foi utilizado uma normalização na matriz de valores singulares ([WANG, 2017](#)). Por fim, o cálculo realizado é descrito no código abaixo:

```

stock_weights_svd = s

top_n = 28
bool_top_stocks = abs(Vh[0].T) > np.sort(abs(Vh[0].T))[len(
    stock_weights_svd)-top_n]
index_fund_weights = (stock_weights_svd[bool_top_stocks])
index_fund_tickers = returns.columns[bool_top_stocks]

```

4.6 RESULTADOS

4.6.1 CONSIDERAÇÕES ESSENCIAIS

O mercado de capitais brasileiro ajusta os seus índices e os preços das ações no decorrer em que é realizada a distribuição de proventos pelas empresas. Os ativos ao distribuírem proventos têm o seu preço final ajustado na data "Ex-provento", o provento é descontado de toda a série histórica do ativo, resultando em um histórico ajustado. O Ibovespa, por sua vez, possui o tratamento de distribuir os proventos em toda a cesta de ações de maneira ponderada, inclusive no próprio ativo pagador de proventos (TOMAZ *et al.*, 2017). Com base nisto, o estudo foi utilizado na série de retornos dos ativos presentes na carteira do Ibovespa, sem serem reajustados retroativamente, utilizando apenas o preço de marcação no fechamento da negociação no dia a ser analisado.



Figura 11 – Comparativo entre o retorno do iBovespa versus o Ibovespa Ponderado

Fonte: Próprio Autor.

Com caráter de eliminar esse efeito, durante a análise, os resultados apresentados serão comparados com a série denominada **Ibovespa Ponderado**, que consiste utilizar a mesma matriz de retornos normalizada, apenas sendo ponderada pela carteira teórica do dia 15 de maio de 2022 (B3, 2022).

Para obtenção do cálculo do retorno ponderado, foi utilizado o código abaixo. Foi utilizando da variável genérica *novo_ibov* que será substituída pela carteira calculada nos métodos de redução de dimensionalidade apresentados.

```

for i, row in returns.iterrows():
    for x in returns.columns:
        if x in index_fund_tickers:
            returns.loc[i,x] = returns.loc[i,x] * novo_ibov[novo_ibov['
                Acoes']==x]['Peso'].
                values[0]

        else:
            returns.loc[i,x] = None
aux = returns
aux = aux.sum(axis=1)+1
aux =aux.to_frame().rename(columns={0:'retorno'})
## Valor inicial em reais para aplica o no primeiro dia de estudo
aux['RESULTADO']=1

df1= aux
for n,(index, row) in enumerate(df1.iterrows()):
    if n ==0:
        df1.loc[index, 'RESULTADO'] = 1
        print(df.loc[index, 'RESULTADO'])
    else:
        df1.loc[index, 'RESULTADO']= df1.iloc[n-1]['RESULTADO']*df1.iloc[
            n]['retorno']

```

O *df1* é *DataFrame* que contém o retorno acumulado, com base na aplicação de R\$ 1,00 no primeiro dia de análise, acruando o retorno diário no decorrer do tempo (Figura 11). O método descrito foi utilizado também para realizar a exibição dos gráficos de retorno dos resultados das carteiras obtidas, com as aplicações das técnicas de redução nas subseções abaixo.

4.6.2 MÉTODO 1

Após a implementação do método 1 com a decomposição em 65 componentes principais e utilizando apenas a primeira componente, foi possível obter um portfolio de 73 ações, nas quais os pesos relativos foram obtidos utilizando a [Equação 3.7](#)

Ação	Peso	Ação	Peso
ABEV3	1.1%	FLRY3	1.25%
ALPA4	1.27%	GGBR4	1.62%
AMER3	1.38%	GOAU4	1.57%
AZUL4	1.3%	GOLL4	1.36%
B3SA3	1.88%	HAPV3	1.16%
BBAS3	1.83%	HYPE3	1.32%
BBDC3	1.78%	IGTI11	1.57%
BBDC4	1.88%	IRBR3	1.32%
BBSE3	1.08%	ITSA4	1.76%
BPAC11	1.99%	ITUB4	1.62%
BPAN4	1.35%	JHSF3	1.77%
BRAP4	1.34%	LREN3	1.72%
BRML3	1.38%	MGLU3	1.12%
CASH3	1.49%	MRVE3	1.62%
CCRO3	1.41%	MULT3	1.61%
CIEL3	1.41%	NTCO3	1.31%
CMIG4	1.53%	PCAR3	1.07%
CMIN3	1.66%	PETZ3	1.0%
COGN3	1.55%	POSI3	1.76%
CPFE3	1.39%	QUAL3	1.14%
CPLE6	1.66%	RAIL3	1.18%
CRFB3	1.24%	RDOR3	1.39%
CSAN3	1.64%	RENT3	1.82%
CSNA3	1.47%	SANB11	1.6%
CVCB3	1.62%	SBSP3	1.27%
CYRE3	1.5%	SOMA3	1.68%
DXCO3	1.56%	SULA11	1.44%
ECOR3	1.84%	TIMS3	1.64%
ELET3	1.15%	TOTS3	1.47%
ELET6	1.16%	UGPA3	1.6%
EMBR3	1.11%	USIM5	1.84%
ENBR3	1.23%	VALE3	1.29%
ENEV3	1.31%	VBBR3	1.42%
ENGI11	1.13%	VIIA3	1.33%
EQTL3	1.51%	VIVT3	1.34%
EZTC3	1.29%	YDUQ3	1.33%

Tabela 2 – Composição da Carteira Teórica Obtida pelo Metodo 1

Fonte: Próprio Autor.

Sendo visível uma melhor distribuição dos ativos em pesos de maneira equilibrada (Figura 12), elimina-se a concentração apresentada na Figura 2

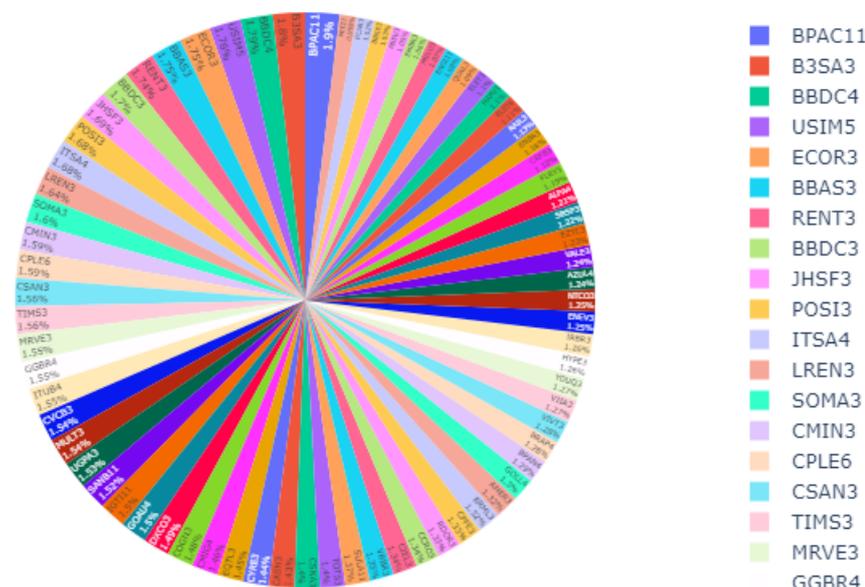


Figura 12 – Gráfico pizza da Carteira Teórica Obtida pelo PCA - Metodo 1

Fonte: Próprio Autor.

É de extrema importância analisar o retorno do portfólio reduzido, que será chamado de **PCA1 Portfólio** em comparação com o portfólio real do Ibovespa, que é composto de 90 ações. Para isso foi realizado o cálculo do retorno de ambos os portfólios, com uma aplicação de R\$ 1,00 no primeiro dia de análise, o resultado é exibido na [Figura 13](#), e é notório uma replicação de comportamentos, que um problema de estabilização de volatilidade, devendo ser discutido na [Capítulo 5](#).

Por fim, é possível afirmar que o método 1 foi capaz de replicar o comportamento do Índice bovespa com uma redução total de 18,89% no tamanho nominal da carteira, além de redistribuir os pesos das ações, mitigando os riscos presentes em decorrer da concentração.

4.6.3 MÉTODO 2

À teste de resultados do método 2, foi utilizado a redução em 10 componentes principais e mantida a abordagem do primeiro método em utilizar apenas a componente principal para a obtenção dos pesos relativos do portfólio com base na [Equação 3.7](#). Foi resultante uma carteira teórica de 60 ativos denominada de **PCA2 Portfólio**, sua composição é exibida na [subseção 4.6.3](#), e sua disposição ponderada pode ser visualizada

pelo gráfico pizza apresentado em [Figura 14](#)

Ação	Peso	Ação	Peso
AZUL4	2.44%	ITSA4	1.13%
B3SA3	1.28%	ITUB4	1.18%
BBAS3	1.23%	JHSF3	1.49%
BBDC3	1.27%	KLBN11	1.2%
BBDC4	1.27%	LWSA3	1.0%
BEEF3	1.19%	MRFG3	1.06%
BPAC11	1.6%	MRVE3	1.34%
BPAN4	1.42%	MULT3	1.09%
BRAP4	2.1%	PCAR3	1.63%
BRKM5	2.43%	PETR3	1.83%
BRML3	1.04%	PETR4	1.63%
CASH3	1.68%	POSI3	2.14%
CCRO3	1.18%	PRIO3	2.61%
CIEL3	1.03%	RAIL3	1.2%
CMIG4	1.02%	RDOR3	1.23%
CMIN3	2.6%	RENT3	1.3%
CPFE3	1.09%	RRRP3	3.04%
CRFB3	1.09%	SANB11	1.18%
CSAN3	1.87%	SLCE3	2.07%
CSNA3	3.63%	SOMA3	1.07%
CVCB3	1.12%	SULA11	1.07%
DXCO3	1.6%	SUZB3	1.53%
ECOR3	1.64%	TOTS3	1.67%
EMBR3	1.64%	UGPA3	1.64%
GGBR4	2.97%	USIM5	2.63%
GOAU4	2.84%	VALE3	2.04%
GOLL4	3.19%	VIIA3	1.13%
HYPE3	1.04%	WEGE3	1.15%
IGTI11	1.0%	YDUQ3	1.47%
IRBR3	1.88%		

Tabela 3 – Composição da Carteira Teórica Obtida pelo PCA - Metodo 2

Fonte: Próprio Autor.



Figura 15 – Comparação do retorno da carteira do Ibovespa e o portfolio obtido com a aplicação do PCA - método 2

Fonte: Próprio Autor.

4.7 DECOMPOSIÇÃO EM VALORES SINGULARES

Utilizando a técnica de decomposição em valores singulares (SVD), foi possível realizar uma redução de 68,89% na carteira do Ibovespa, possuindo uma carteira final de 28 ações, que pode ser vista na [Tabela 4](#), e ilustrada em forma de gráfico pizza na [Figura 16](#)

Ação	Peso
ASAI3	14.11
BBSE3	8.69
EMBR3	2.11
ENBR3	2.06
ENEV3	1.99
BEEF3	8.32
CMIN3	4.91
COGN3	4.39
CPFE3	4.27
BRAP4	7.21
EQTL3	1.60
FLRY3	1.40
GGBR4	1.35
GOAU4	1.19
IGTI11	0.92
LREN3	0.46
POSI3	0.11
PRIO3	0.06
BRKM5	6.58
CMIG4	5.11
CPLE6	4.00
CSAN3	3.70
CSNA3	3.23
CYRE3	2.97
EGIE3	2.74
ELET3	2.50
ELET6	2.27
ENGI11	1.75

Tabela 4 – Carteira Reduzida em 28 ações com o método de Decomposição em Valores Singulares

Fonte: Próprio Autor.

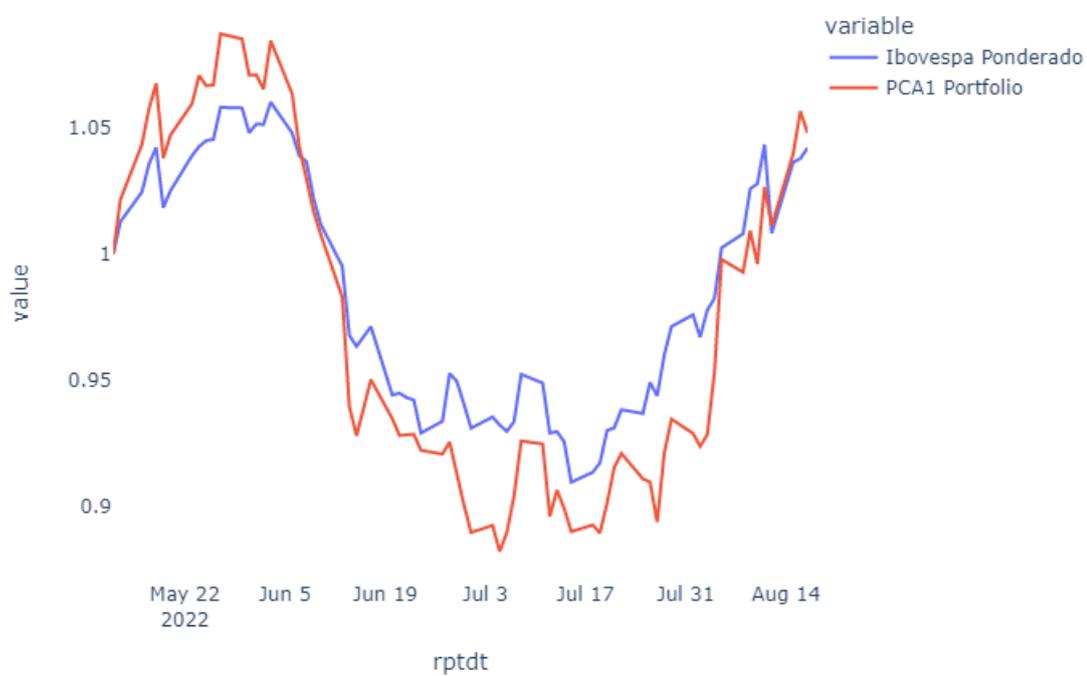


Figura 13 – Comparação do retorno da carteira do Ibovespa e o portfolio obtido com a aplicação do PCA - método 1

Fonte: Próprio Autor.

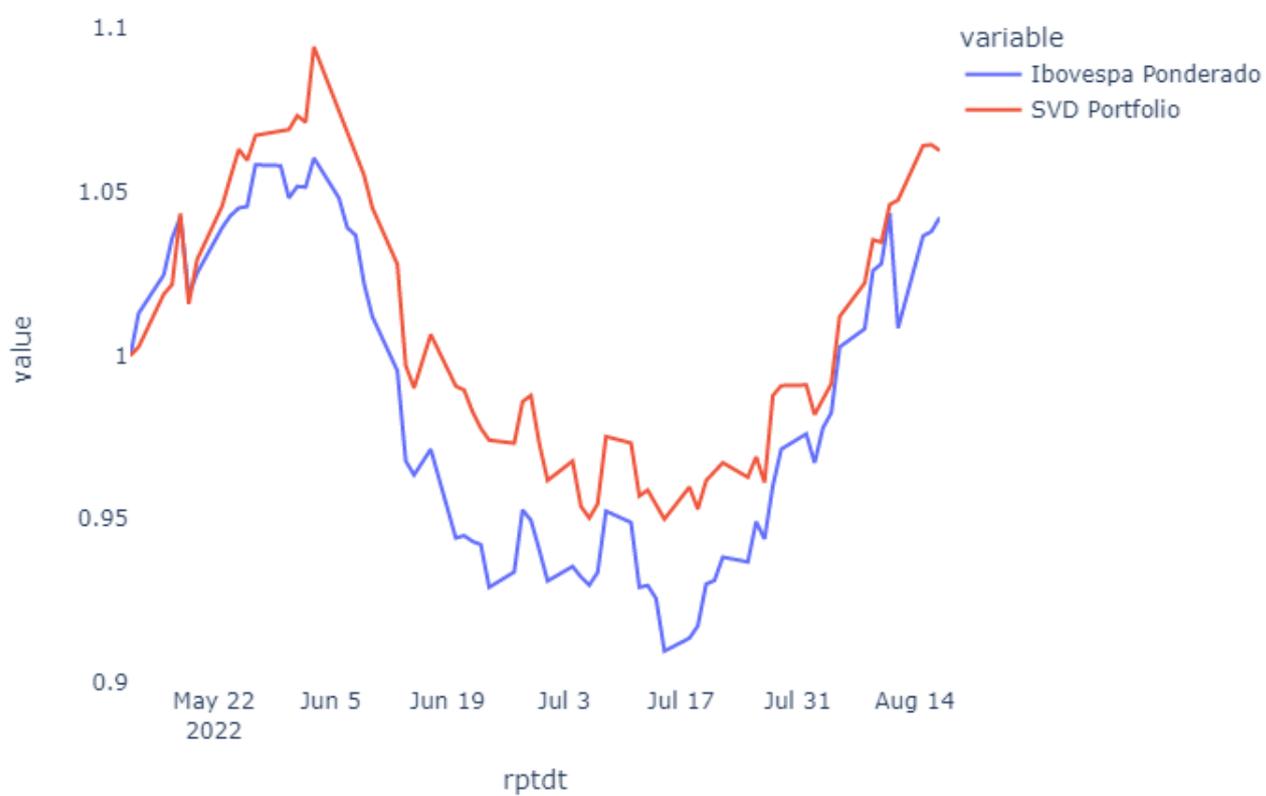


Figura 17 – Retorno observado do SVD Portfólio ao lado do retorno do Ibovespa Ponderado

Fonte: Próprio Autor.

5 CONCLUSÃO

Foi possível cumprir os objetivos específicos apresentados, visto que, o intuito desse trabalho era de buscar uma verificação a aplicabilidade das técnicas de redução de dimensionalidade para replicar o comportamento do Ibovespa.

Utilizando a análise de redução de componentes principais, em ambos os métodos apresentados, foi possível obter uma redução de 18,89% (Método 1) e 33,33% (Método 2). Além disso, os métodos resultaram portfólios mais equilibrados, eliminando os riscos de uma concentração de ativos, como é perceptível na carteira original do Índice Bovespa, no qual 10 ações representam mais que 50% do peso final da alocação de capital.

Em contrapartida, ao utilizar a técnica de decomposição em valores singulares, foi notório uma redução mais agressiva de 68,89%. Devido a quantidade de ações presentes no índice de ações brasileiro não ser elevada, e ademais, concentrada, uma redução significativa acaba comprometendo a fidelidade que era esperada ao tentar realizar uma replicação.

Por fim, é perceptível que ao utilizar apenas os métodos de redução de dimensionalidade, não é possível tornar o retorno diário do portfólio reduzido ajustado ao retorno diário do Ibovespa, conseguindo apenas fidelizar o seu comportamento. Para isso, é preciso inserir uma abordagem para solução de um problema de controle, podendo ser aplicado controladores ou algoritmos de aprendizado de máquina, com objetivo de direcionar o retorno do portfólio para o retorno de referência.

Neste trabalho foi apresentado um experimento utilizando a carteira vigente do iBovespa, visto que, devido a rotatividade trimestral, componentes pertencentes as carteiras passadas seriam incorporados à análise de carteiras vigentes. Sendo assim, para uma validação do algoritmo em uma série histórica, foi interpretado que caso seja necessário, é preciso realizar o estudo por blocos, analisando cada série vigente de carteiras de forma separada, como abordado no trabalho desenvolvido, apenas deslocando o modelo em janelas passadas.

REFERÊNCIAS

- ALDRIDGE, I.; AVELLANEDA, M. *Big data science in finance*. [S.l.]: John Wiley & Sons, 2021. Citado 3 vezes nas páginas 20, 27 e 30.
- B3. *Composição da carteira*. 2022. <https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carreira.htm>. [Accessed 27-Aug-2022]. Citado na página 31.
- BOLDRINI, J. L. et al. *Álgebra linear*. [S.l.]: Harper & Row, 1980. Citado na página 15.
- DUNN, K. Process improvement using data. *Experimentation for Improvement*. Hamilton, Ontario, Canada. *Creative Commons Attribution-ShareAlike*, 2019. Citado na página 19.
- GEISS, J. Latent semantic indexing and information retrieval-a quest with bosse. 2006. Citado na página 22.
- MONTGOMERY, D. C.; RUNGER, G. C.; CALADO, V. *Estatística Aplicada E Probabilidade Para Engenheiros*. [S.l.]: Grupo Gen-LTC, 2000. Citado 2 vezes nas páginas 15 e 16.
- RUDY, S. et al. Data-driven identification of parametric partial differential equations. *SIAM Journal on Applied Dynamical Systems*, SIAM, v. 18, n. 2, p. 643–660, 2019. Citado 3 vezes nas páginas 20, 21 e 22.
- SHLENS, J. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014. Citado na página 17.
- TOMAZ, N. C. B. et al. Relação entre a distribuição de dividendos e o preço da ação das empresas listadas na b3. Universidade Federal da Paraíba, 2017. Citado na página 31.
- VARELLA, C. A. A. Análise de componentes principais. *Seropédica: Universidade Federal Rural do Rio de Janeiro*, p. 38, 2008. Citado 2 vezes nas páginas 17 e 18.
- VIEIRA, D. *Tipos de Ações e Seus Códigos de Negociação* — *dvinvest.com.br*. 2017. <<https://dvinvest.com.br/aprenda/blog/tipos-de-acoes-e-seus-codigos-de-negociacao>>. [Accessed 29-Aug-2022]. Citado na página 12.
- WANG, D. Adjustable robust singular value decomposition: Design, analysis and application to finance. *Data*, MDPI, v. 2, n. 3, p. 29, 2017. Citado na página 30.
- YANG, L. An application of principal component analysis to stock portfolio management. University of Canterbury. Department of economics and finance, 2015. Citado na página 28.

Anexos

ANEXO A – CORRELAÇÃO SETORIAL

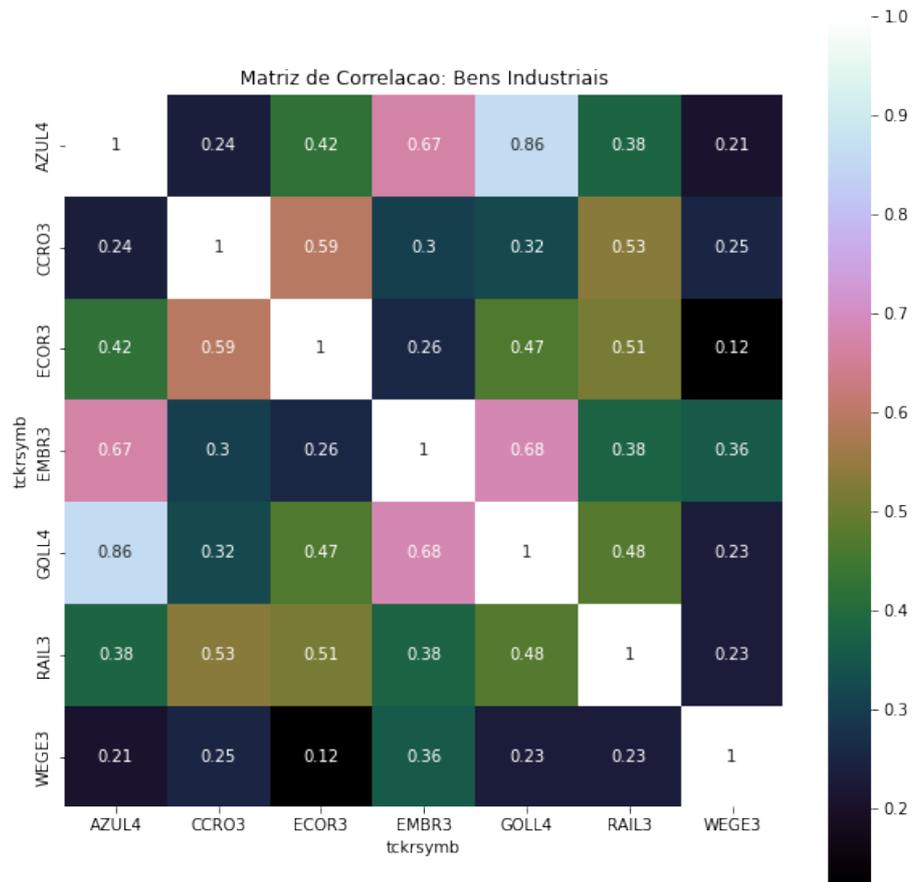


Figura 18 – Matriz de correlação: Bens Industriais

Fonte: Próprio Autor.

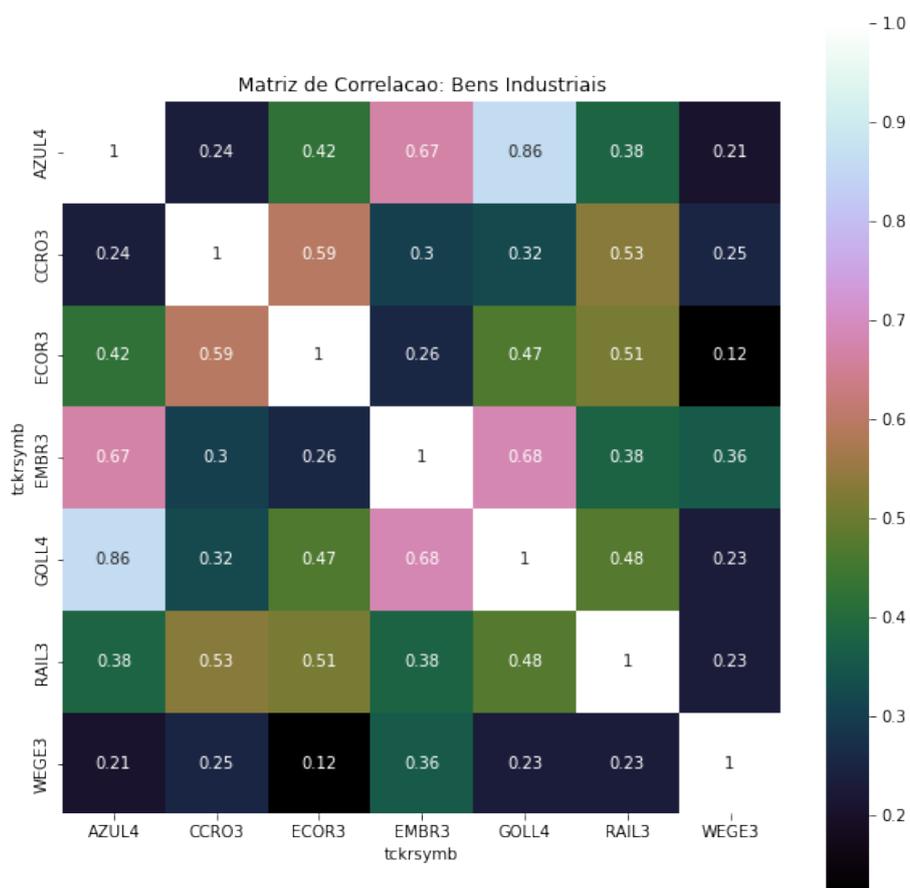


Figura 19 – Matriz de correlação: Bens Industriais

Fonte: Próprio Autor.

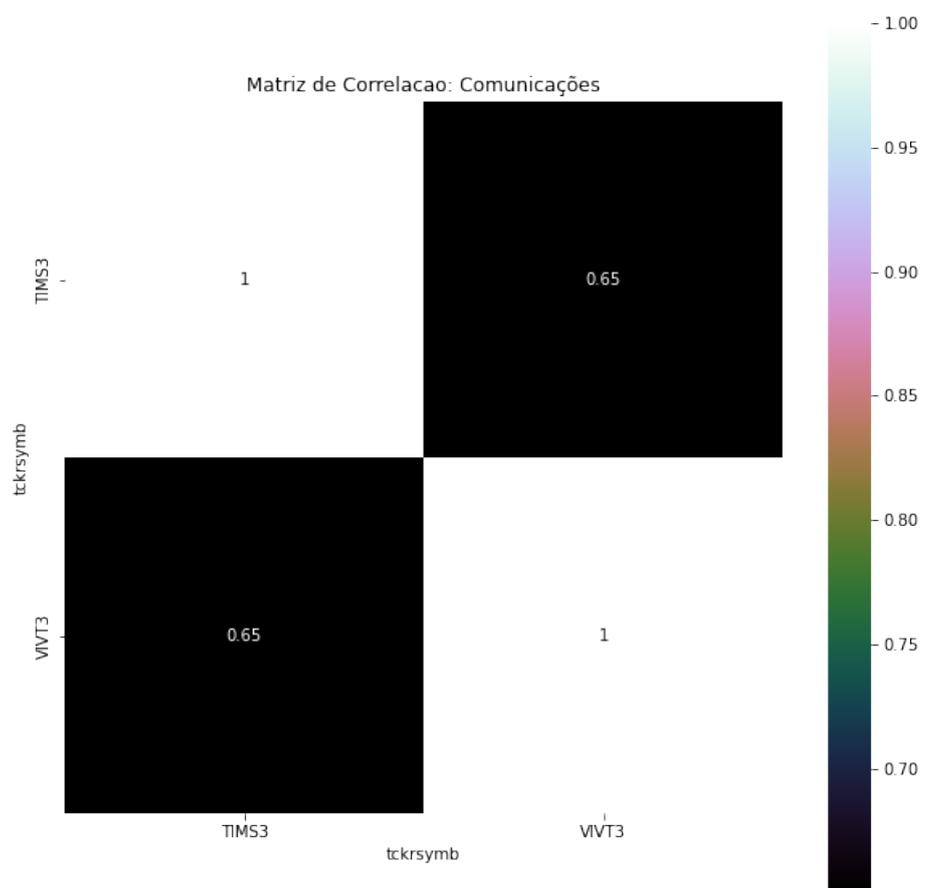


Figura 20 – Matriz de correlação: Comunicações

Fonte: Próprio Autor.

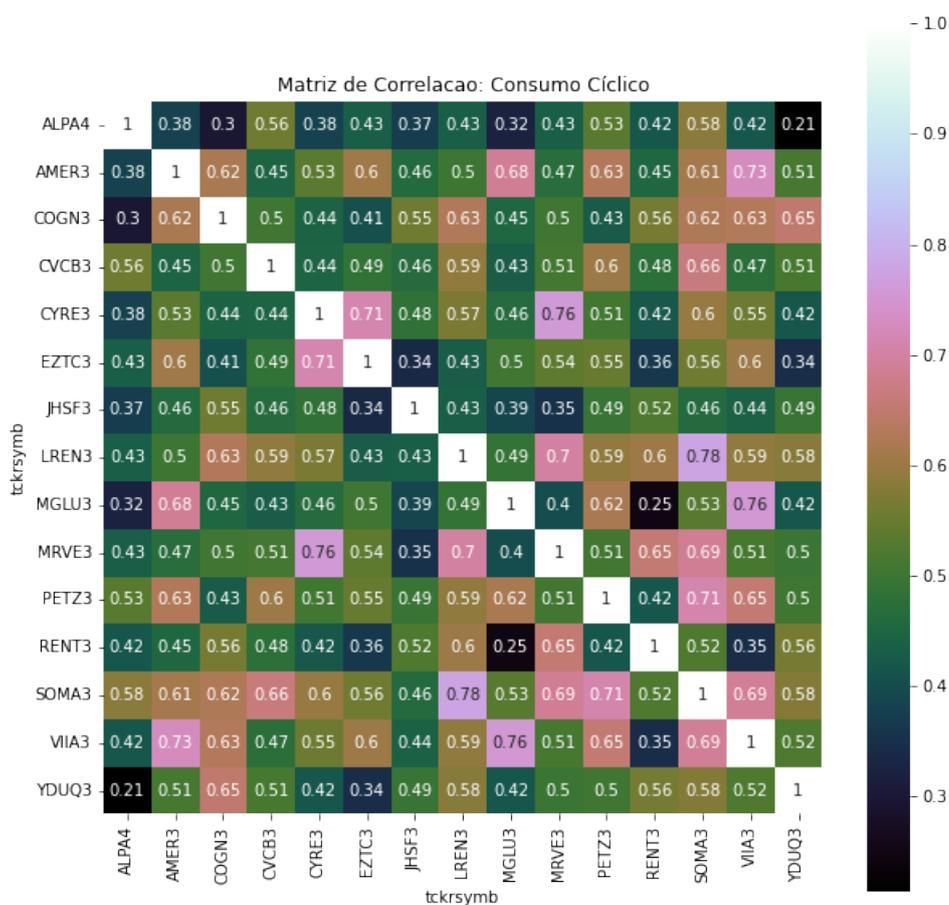


Figura 21 – Matriz de correlação: Consumo Cíclico

Fonte: Próprio Autor.

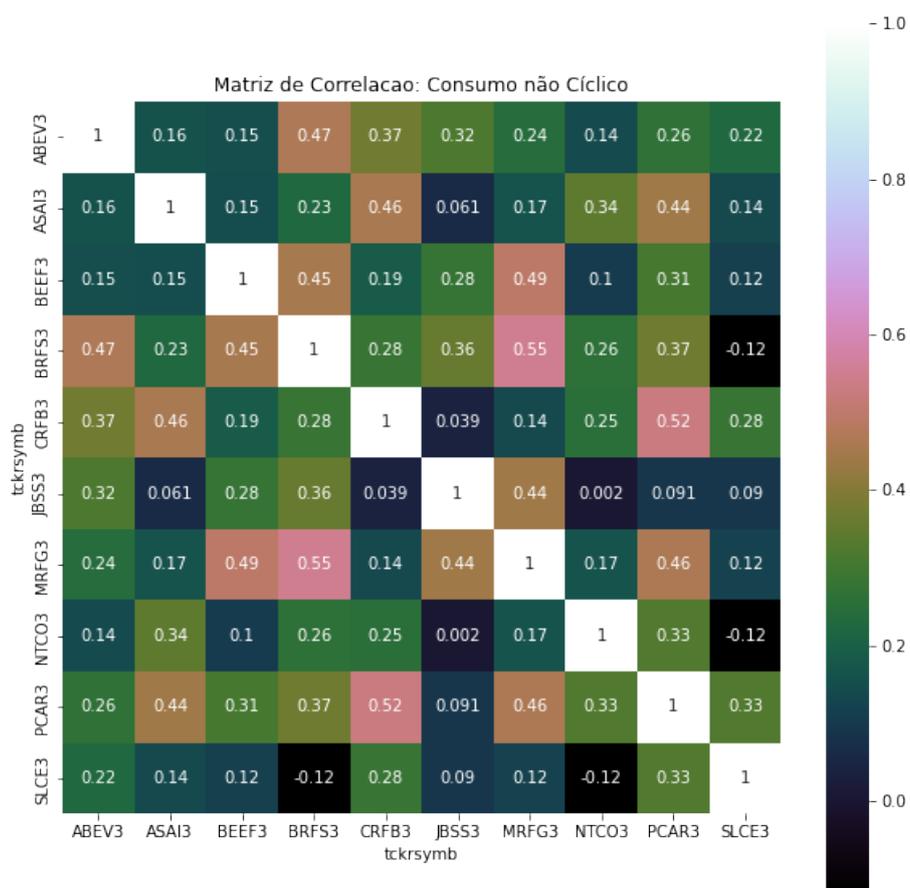


Figura 22 – Matriz de correlação: Consumo Cíclico

Fonte: Próprio Autor.

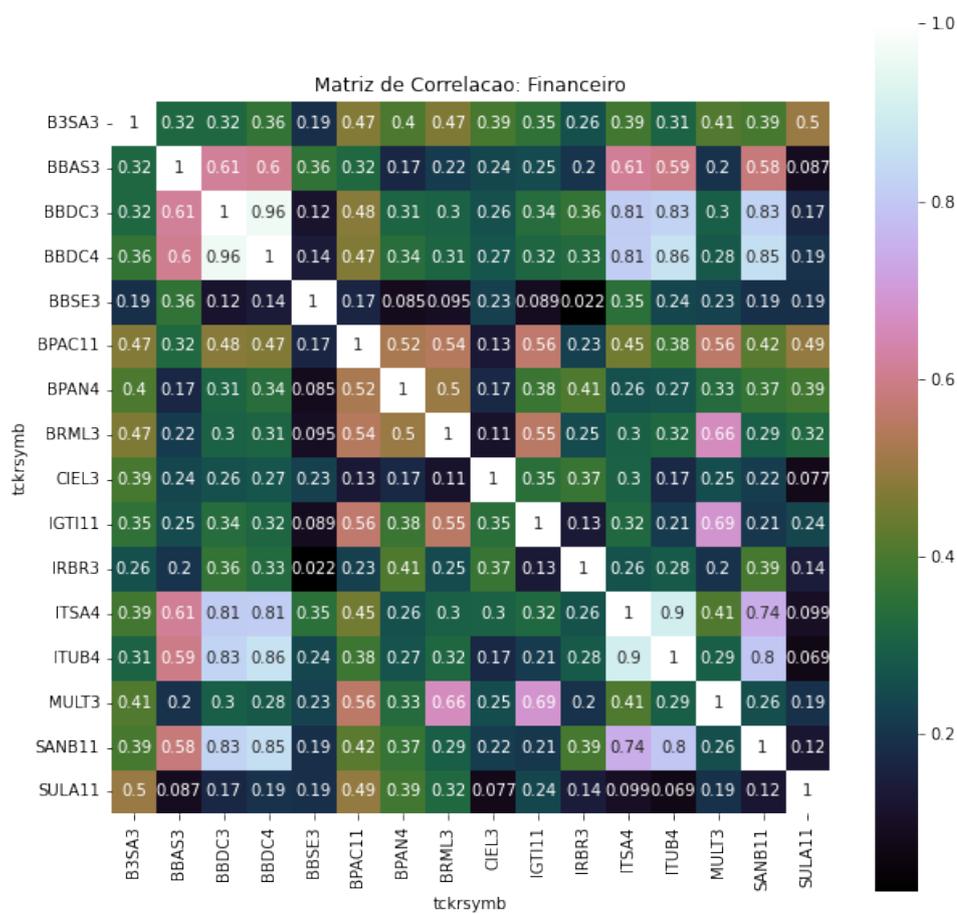


Figura 23 – Matriz de correlação: Financeiro

Fonte: Próprio Autor.

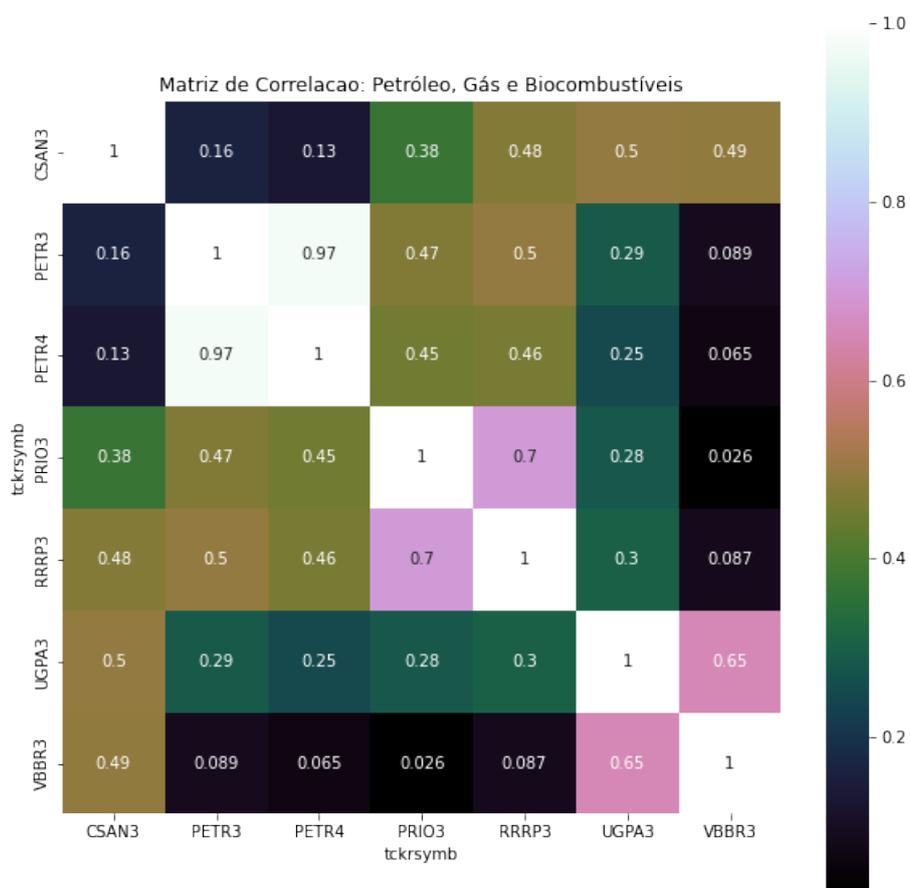


Figura 24 – Matriz de correlação: Petróleo, Gás e Biocombustíveis

Fonte: Próprio Autor.

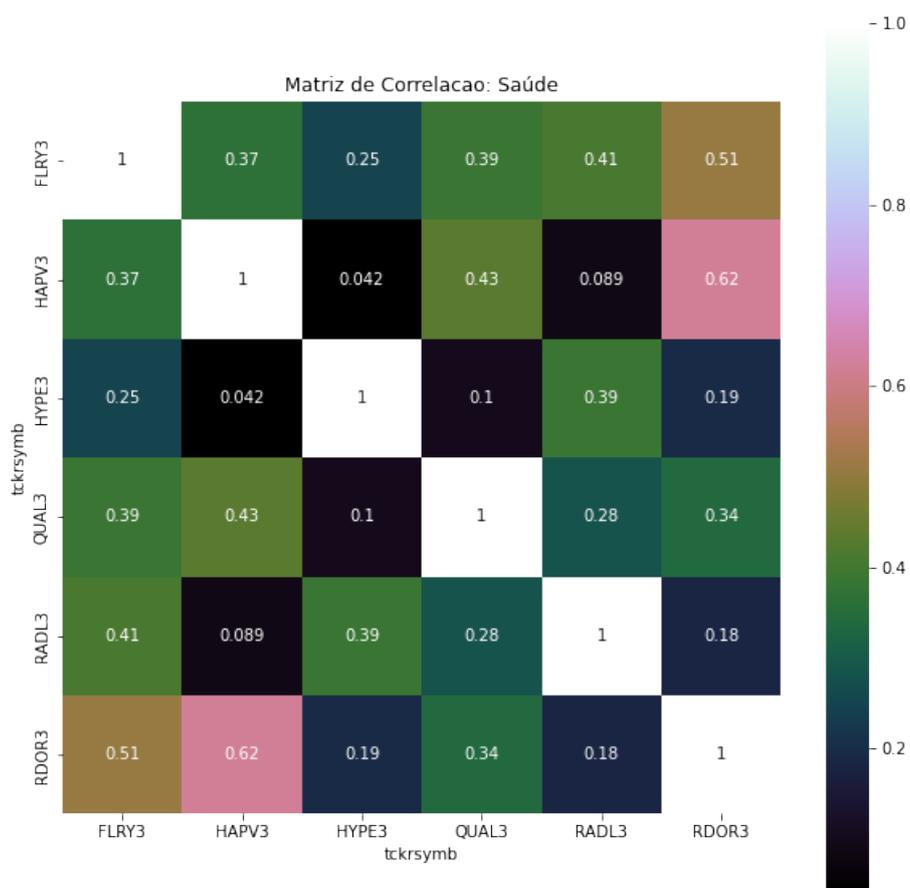


Figura 25 – Matriz de correlação: Saúde

Fonte: Próprio Autor.

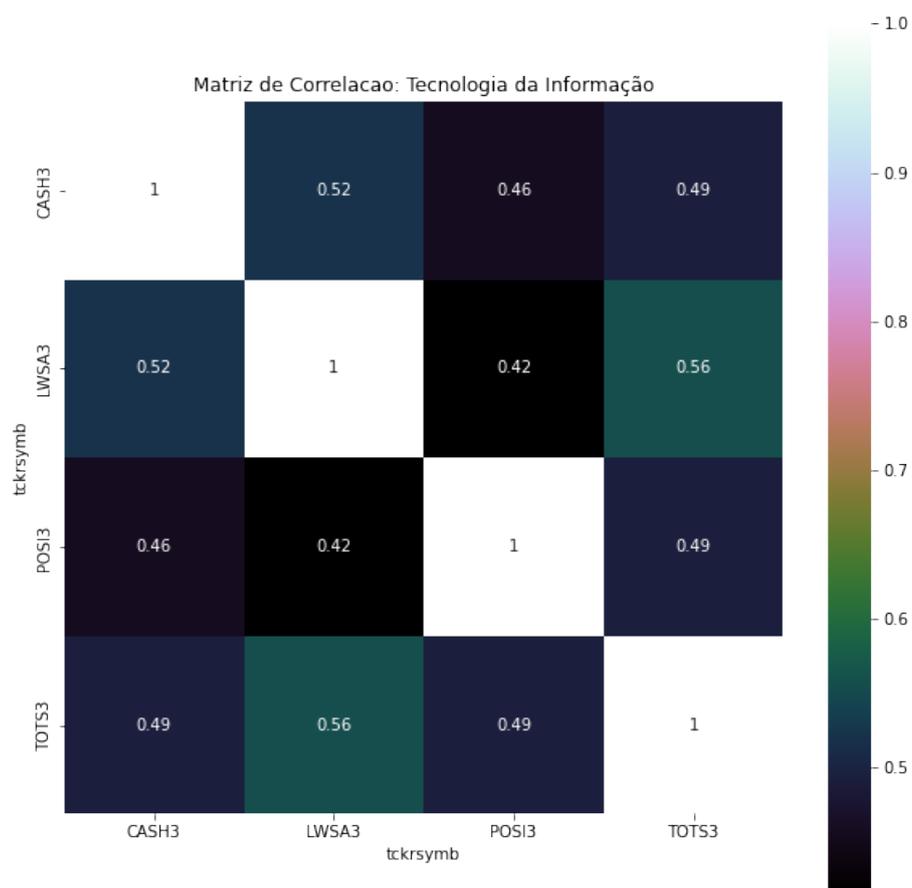


Figura 26 – Matriz de correlação: Tecnologia da Informação

Fonte: Próprio Autor.