

Estudo de Variáveis Econômicas para a Previsão de Tendência do Mercado de Ações Via Aprendizado de Máquina

Projeto de Iniciação Científica

Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP

Proc. no. 2024/04789-2

Relatório Final

Período: 01/07/2024 a 24/04/2025

Estudante: Leonardo Rodrigues da Silva

l251773@dac.unicamp.br

Orientador: Prof. Dr. Guilherme Palermo Coelho

guilherme@ft.unicamp.br

Limeira, 24 de abril de 2025

Informações gerais:

Título do Projeto: Estudo de Variáveis Econômicas para a Previsão de Tendência do Mercado de Ações Via Aprendizado de Máquina

Processo: 2024/04789-2

Vigência: 01/07/2024 a 30/06/2025

Estudante beneficiário: Leonardo Rodrigues da Silva

Orientador: Prof. Dr. Guilherme Palermo Coelho

Instituição Sede: Faculdade de Tecnologia (FT), Universidade Estadual de Campinas (Unicamp).

Resumo

O mercado de ações tem crescido nos últimos anos, de modo que cada vez mais pessoas estão começando a investir seu dinheiro em diversos produtos existentes nas corretoras financeiras. Segundo dados disponibilizados pela Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA), em parceria com o DataFolha, o percentual de investidores brasileiros passou de 31% da população, em 2021, para 36%, em 2022, correspondendo a um acréscimo de 8 milhões de pessoas. Diante desse cenário, é importante compreender melhor como funcionam as oscilações do mercado financeiro para que se tenha uma maior previsibilidade de seus movimentos, aumentando a segurança dos investimentos. Nesse contexto, o presente projeto busca contribuir para essa maior compreensão do mercado de ações, tentando identificar se existem variáveis econômicas que, quando combinadas a dados técnicos históricos de uma empresa, melhoram a capacidade de predição da tendência de comportamento de suas ações, via algoritmos de aprendizado de máquina. Para isso, serão desenvolvidas as seguintes atividades: (i) análise aprofundada de um determinado setor econômico escolhido para estudo; (ii) seleção de indicadores macro/macroeconômicos relevantes que podem influenciar a variação dos preços de ações; (iii) buscar encontrar a ligação entre os atributos selecionados e os valores de fechamento da ação; e (iv) criação de um modelo que realiza a previsão de tendências com base nos atributos selecionados e cotações históricas da ação. Nesta primeira etapa do projeto, foram realizadas as atividades iniciais de construção de um modelo LSTM (*Long Short Term Memory*) para identificar a melhor forma de prever a tendência de uma determinada ação apenas com os preços de fechamento ajustado, o início dos estudos sobre indicadores econômicos e sua relevância para determinado ativo e compreensão da série temporal do mesmo e como extrair da melhor forma informação de seus valores para melhorar a previsão.

Palavras-chave: Previsão de ações, Aprendizado de Máquina, Mercado Financeiro, Redes Neurais.

Abstract

The stock market has grown in recent years, with more and more people starting to invest their money in various products available at financial brokers. According to data provided by the Brazilian Association of Financial and Capital Market Entities (ANBIMA), in partnership with DataFolha, the percentage of Brazilian investors increased from 31% of the population in 2021 to 36% in 2022, corresponding to an addition of 8 million people. In this context, it is important to better understand how the financial market fluctuations work in order to have a greater predictability of its movements, increasing investment security. In this regard, the present project seeks to contribute to a better understanding of the stock market by trying to identify whether there are economic variables that, when combined with historical technical data of a company, improve the ability to predict the trend of its stock behavior through machine learning algorithms. The following activities will be developed: (i) in-depth analysis of a specific economic sector chosen for study; (ii) selection of relevant macro/microeconomic indicators that may influence stock price variation; (iii) finding the connection between the selected attributes and the stock closing values; and (iv) creation of a model that forecasts trends based on the selected attributes and historical stock quotes. In the first phase of the project, initial activities were carried out to build an LSTM (*Long Short Term Memory*) model to conduct experiments to identify the best way to predict the trend of a given stock using only adjusted closing prices. Additionally, studies on economic indicators were initiated to assess their relevance to a specific asset, along with an understanding of the time series and how to optimize its values to improve predictions.

Keywords: Stock market, machine learning, trend prediction.

Sumário

Resumo	2
Abstract	3
Introdução e Justificativa	5
Objetivos	7
Fundamentação teórica: Indicadores Macro/Microeconômicos	8
Metodologia	10
Cronograma e próximos passos	23
Referências Bibliográficas	25

1. Introdução e Justificativa

A bolsa de valores brasileira, [B]³, é onde diversas empresas de capital aberto disponibilizam frações de seu patrimônio (as ações) para aquisição por pessoas físicas e jurídicas. Em um processo benéfico para ambas as partes, a empresa recebe recursos para promover expansões e financiar grandes investimentos e o acionista participa das distribuições de lucros e pode valorizar seu patrimônio com o crescimento da empresa. Segundo dados disponibilizados pela Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA), em parceria com o DataFolha, o percentual de brasileiros investidores no mercado de ações passou de 31% da população, em 2021, para 36%, em 2022, correspondendo a um acréscimo de 8 milhões de pessoas (ANBIMA, 2023). Os dados do *Raio X do Investidor Brasileiro* apontam que o maior impulso para esse crescimento veio da classe C. Nesse contexto, é importante buscar compreender melhor o mercado financeiro, tornando-o mais acessível a todos os estratos sociais existentes, o que pode ser feito através do desenvolvimento de algoritmos que garantem maior segurança na escolha dos ativos que receberão investimentos. E nesse sentido, de acordo com a Comissão de Valores Mobiliários (CVM, 2004), “*está comprovado que as pessoas, até mesmo aquelas com poucos recursos, quando se empenham nessa jornada, alcançam a segurança financeira almejada*”. Desse modo, é necessário disponibilizar ferramentas que ajudem na tomada de decisão dos investidores e que minimizem suas perdas.

Atualmente, os profissionais da área de investimentos e mercado de ações usam dois métodos para tentar prever o movimento do mercado: a análise *técnica* e a *fundamentalista*. A análise técnica surgiu no meio profissional do mercado financeiro e seus adeptos são, majoritariamente, especuladores no mercado. Já a análise fundamentalista surgiu no meio acadêmico e seus adeptos são administradores de fundos e investidores a longo prazo (Barros, 2015). Enquanto a análise técnica procura prever os movimentos futuros com base em gráficos que representam valores históricos de fechamento de preço e volume de negociação, dentre outros, a análise fundamentalista busca compreender a situação financeira da

empresa como um todo, analisando seus demonstrativos financeiros, fluxo de caixa, *dividend yield* e balanço patrimonial, além de suas perspectivas, com base nas suas políticas internas, com a finalidade obter uma visão ampla sobre as alternativas de investimentos. Resumidamente, a análise fundamentalista tenta entender o negócio para obter resultados no médio e longo prazo. De acordo com Assaf Neto (2014), “a premissa básica da avaliação fundamentalista é obter um valor justo, o valor reflete o retorno esperado baseado em projeções de desempenho futuro coerentes com a realidade do negócio em avaliação.” Ou seja, o foco da análise fundamentalista é olhar diversas variáveis com a finalidade de se chegar em um preço justo na atual conjectura da empresa. Entretanto, é nítido que ambas as estratégias demandam bastante tempo e esforço para serem executadas com eficiência, e mesmo assim, podem gerar resultados errôneos.

Segundo a Hipótese do Mercado Eficiente (Fama, 1970), um mercado onde os preços refletem todas as informações existentes é chamado mercado eficiente. Com base no comportamento do mercado, é possível identificar três formas de eficiência: fraca, semiforte e forte. A forma fraca indica que os preços refletem toda a informação contida nos registros dos preços passados. A segunda forma (semiforte), defende que os preços refletem, além dos registros passados de preços, todas as informações publicadas relacionadas à empresa. A forte alega que os preços refletem todos os critérios citados anteriormente com o adicional de informações privadas, cujo acesso não é público. A partir disso, é possível concluir que a análise técnica está mais alinhada com a teoria da eficiência fraca, já que ela tenta prever os preços futuros com as informações existentes nos preços, como volume de negociações e preço de fechamento. Por outro lado, a análise fundamentalista tenta extrair informações nos indicadores macro/microeconômicos, verificando a saúde da empresa. Portanto, ambas as análises possuem abordagens válidas e deveriam ser combinadas para melhorar a qualidade da predição.

Apesar da variação dos preços das ações aparentarem ter um comportamento aleatório ao longo do tempo, existem meios de analisar a série temporal e tentar prever o movimento com base no comportamento passado. Para

ilustrar, em sua dissertação, Mueller (1996) buscou verificar se uma rede neural artificial, utilizando o algoritmo de *backpropagation* para treinamento, tem um desempenho melhor na predição em comparação com os métodos convencionais de previsão de séries temporais, dentre eles média móvel, alisamento exponencial linear, modelo autorregressivo, método de Box e Jenkins etc. A série temporal usada nos experimentos foi o preço de fechamento diário de um conjunto de ações ordinárias de empresas nacionais, coletadas entre 1986 e 1990. Na conclusão do trabalho, foi observado que as redes neurais artificiais apresentaram desvios de menor magnitude quando comparadas aos métodos tradicionais, reforçando o potencial dessa ferramenta na identificação de padrões em séries temporais, focada no mercado acionário.

Diante disso, neste trabalho será definido um setor de atuação econômico e escolhidas ações de empresas relevantes desse setor, negociadas na bolsa de valores brasileira ([B]³). Feito isso, serão coletados os dados históricos de negociação das ações de cada empresa, como seus valores diários de fechamento, para que possam ser utilizados no modelo proposto de predição. Após essa coleta, será feito um mapeamento, a partir do setor das empresas, com a finalidade de encontrar indicadores-chave que podem influenciar diretamente na oscilação da cotação de suas ações. Coletados esses dados, eles serão utilizados em modelos baseados em aprendizado de máquina, construídos com o intuito de indicar a tendência do movimento de cada ação.

2. Objetivos

O objetivo deste projeto de iniciação científica é o estudo da influência dos indicadores macro e microeconômicos relacionados a um setor econômico nos preços das ações de empresas que atuam em tal setor. Para atingir esse fim, o projeto possui as seguintes metas específicas:

- Realização de um levantamento bibliográfico sobre trabalhos que abordam a influência de indicadores macro e microeconômicos nos preços de ações.
- Escolher algumas empresas que se destaquem em um setor econômico (*Commodities*) como casos de estudo, de modo que a influência de tais indicadores possa ser efetivamente verificada.
- Identificação e seleção desses indicadores.
- Realização de um estudo comparativo entre modelos de predição baseados em LSTM que utilizem tanto dados combinados quanto apenas valores históricos de cotação.
- Proposição de uma estratégia de investimento simples, que possa se beneficiar das predições geradas pelos modelos de predição criados, e comparar também o desempenho dos modelos com base em simulações de investimento.

3. Fundamentação teórica

3.1 Indicadores Macro/Microeconômicos

Normalmente, as análises técnicas ou fundamentalistas focam nos dados que têm ligação direta com a variação das ações, como preço de fechamento, volume de negociação ou preço de fechamento ajustado. Entretanto, nem sempre tais análises exploram o uso tanto de indicadores macroeconômicos quanto microeconômicos para potencializar a previsão de ter uma maior qualidade na tomada de decisão.

Primeiramente, é importante esclarecer o que são indicadores macro e microeconômicos. Segundo Bomfim, Macedo e Marques (2019), os indicadores macroeconômicos são índices que reúnem números sobre a situação econômica de determinada região ou setor. Esses dados mostram o desempenho financeiro do objeto em observação. Dessa forma, é lógico concluir que esses indicadores podem ajudar a compreender melhor o movimento do valor da ação, já que eles incorporam informações que podem melhorar a previsão da tendência de uma ação. Alguns

exemplos de indicadores são o Produto Interno Bruto (PIB), Índice de Preços ao Consumidor Amplo (IPCA), Taxa de Juros Básica (SELIC), Índice Nacional de Preços ao Consumidor (INPC), Índices Geral de Preços - Mercado (IGP-M), Balança Comercial, Câmbio, CDI e Taxa Referencial.

De acordo com Ribeiro, Macedo e Marques (2012), os indicadores microeconômicos são aqueles que estão mais relacionados com o desempenho de um determinado nicho ou objeto único. Estes dizem respeito diretamente ao desempenho financeiro de uma empresa. Por exemplo, indicadores microeconômicos relacionados à rentabilidade da empresa são o retorno sobre patrimônio líquido (ROE), retorno sobre investimento (ROI), retorno sobre ativos (ROA). Indicadores relacionados à lucratividade podemos citar margem de lucro bruta, margem de lucro líquida e margem EBITDA (Lucros antes de Juros, Impostos, Depreciação e Amortização). E por último, para os dados relacionados ao endividamento da empresa podem ser mencionados Imobilização dos Recursos a Longo Prazo (IRPL), Participação de Capitais de Terceiro (PCT), Composição do endividamento (CE), Imobilização do Patrimônio Líquido (IPL), Índice de Endividamento Financeiro (EF), Índice de Endividamento Geral (EG).

Assim, esses indicadores dizem muito sobre o desempenho e a saúde financeira de uma empresa. Portanto, é nítido pensar que eles podem fornecer bastante informação sobre a ação de uma empresa e que seu uso também pode ter impacto na qualidade da previsão.

Bomfim, Macedo e Marques (2019) realizaram um estudo para verificar se indicadores financeiros e operacionais têm alguma relação ou influência para a avaliação de desempenho de empresas do setor de petróleo e gás. Foram selecionados dados de 55 empresas do setor, listadas na New York Exchange (NYSE), e foi utilizada a técnica de análise fatorial para entender a influência de 10 indicadores. Com os resultados da pesquisa, foi observado que existem três fatores principais que influenciam o desempenho das empresas observadas: rentabilidade, alavancagem e desempenho operacional. Isso comprova que tais indicadores

microeconômicos possuem valor dentro da previsão de uma ação e seu uso é importante.

Em relação ao uso de indicadores macroeconômicos, Santos (2023) realizou um trabalho para previsão de séries temporais utilizando técnicas fundamentalistas e técnicas econométricas. Ele realizou uma boa revisão da literatura, discutindo sobre a maioria dos indicadores e sua relação com o mercado. A partir da análise dos indicadores em combinação com os demonstrativos financeiros das empresas mais fatos que ocorreram entre 2019 e 2022, como a pandemia de covid-19, Santos (2023) identificou as empresas que obtiveram os melhores desempenhos no setor de *commodities* e depois extraiu os demonstrativos financeiros para prever os preços. Além disso, uma preocupação que o autor teve, que também se aplica ao presente trabalho, é a verificação da série temporal em relação à sua estacionariedade. Uma série estacionária possui características constantes ao longo do tempo. Dessa forma, é importante aplicar testes estatísticos aos dados, tais como teste de Dickey-Fuller aumentado, para garantir que a série seja estacionária.

3.2 Uso de LSTM's em investimentos

O mercado financeiro é um ambiente rico em dados sequenciais, onde o histórico influencia diretamente o futuro. Modelos tradicionais muitas vezes têm dificuldade em capturar dependências de longo prazo presentes nas séries temporais de preços, volumes e indicadores.

Neste contexto, as redes LSTM, uma evolução das Redes Neurais Recorrentes, mostram-se particularmente interessantes. Sua arquitetura é projetada especificamente para aprender e reter informações relevantes ao longo de extensos períodos, superando limitações de "memória" de modelos mais simples. Essa capacidade de conectar eventos passados a movimentos futuros é crucial para analisar a dinâmica complexa dos ativos financeiros.

Por essa razão, a aplicação de LSTMs em tarefas como previsão de tendências, análise de volatilidade e processamento de notícias financeiras tem ganhado tração. O que antes era mais explorado academicamente, hoje se

consolida como uma ferramenta cada vez mais presente na análise quantitativa e no desenvolvimento de estratégias de investimento, como discute Fraga (2021), em sua dissertação sobre integração de redes neurais na otimização de portfólios de investimentos.

Seu trabalho foi feito de uma maneira prática de simular o cotidiano e as atividades exercidas por uma corretora de investimentos, que tem como responsabilidade gerir capitais depositados por pessoas físicas ou jurídicas com a confiança de que essa gestão dará lucro e valorização do patrimônio. Sua metodologia foi realizar uma pré-seleção de ativos elegíveis que fazem parte do portfólio IBRX-50, depois disso previsão do retorno nos próximos 20 dias por LSTMs e previsão da covariância dos próximos 20 dias por modelos ARCH/GARCH e LSTMs e por como última etapa, realiza a otimização do portfólio pelo método de Markowitz. Como resultados, obteve um retorno de 70,94% no período de 1 de março a 26 de dezembro de 2019, totalizando uma volatilidade anual de 18,16%. Esses números ficaram acima da Taxa Selic, que variou 11% ao longo do período todo, do índice Ibovespa e do próprio índice IBRX-50, que tem sua composição reavaliada quadrimestralmente, logo não é igual a composição gerada pelo autor.

Com essa conclusão, é exemplificado que métodos tradicionais de gestão de ativos combinados com novas abordagens e olhares pode produzir excelentes resultados e mostrar novas informações que antes jamais poderiam ser extraídas se não fossem os algoritmos de *machine learning*.

4. Metodologia

A Figura 4 contém o fluxograma de todas as atividades que serão desenvolvidas ao longo do projeto, atividades estas que serão descritas e discutidas nas subseções que se seguem. Na Figura 4, verde indica etapas que já foram realizadas, laranja etapas que estão em andamento e cinza etapas ainda não iniciadas.

4.1 Definição de um setor empresarial

Para a execução do presente trabalho, é crucial definir um setor empresarial específico para ser possível, nas etapas seguintes, realizar um estudo desse setor e buscar entender quais indicadores micro/macroeconômicos influenciam a realidade daquela empresa situada naquele setor empresarial do mercado. Inicialmente, foram separados três setores candidatos para a realização deste estudo devido ao fato de que estes compõem aproximadamente 65% do Ibovespa (Gerbelli, 2023): financeiro, energia e *commodities*.

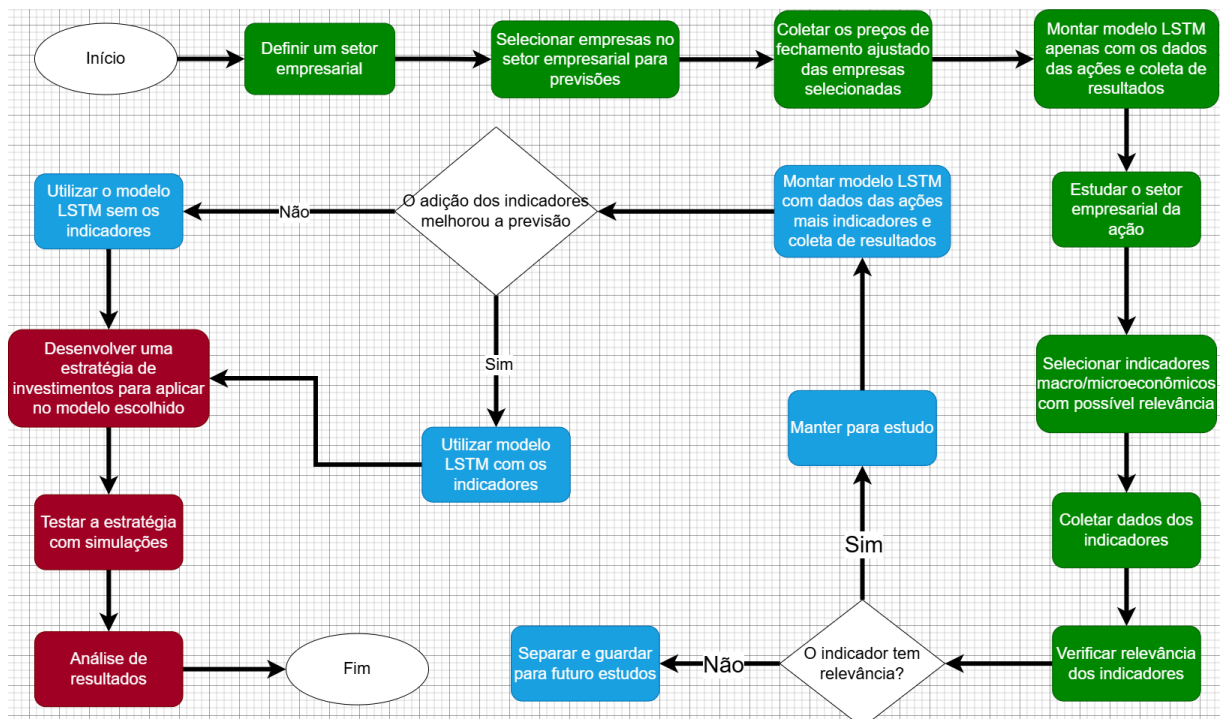


Figura 4: Fluxograma da metodologia do trabalho proposto. Cor verde: Estudo e análise inicial. Cor azul: Execução dos testes sugeridos no estudo. Cor vermelha: Verificação dos resultados na prática.

Apesar de todos estes setores terem um grande impacto na economia brasileira e possuírem bastante indicadores macro e microeconômicos para serem analisados, foi escolhido aqui o setor das *commodities*. Essa escolha foi baseada no fato de que o Brasil é um país que tem uma política de exportação de recursos grande e uma dependência de outros países. Além disso, essa centralização dos investimentos no setor das *commodities* compromete a formação de uma

industrialização diversificada com geração de lucro e acumulação de capital para as economias nacionais (Brandão, Vogt, 2020). Visto isso, prever melhor o comportamento deste setor irá contribuir para o melhor desenvolvimento dessa política e identificar os períodos de baixa para ajudar na escolha de uma estratégia para minimizar os prejuízos.

4.2 Selecionar empresas no setor empresarial para gerar previsões

Como dito anteriormente, foi definido um setor empresarial para ser estudado e analisado com a finalidade de encontrar indicadores que possam influenciar os resultados financeiros das empresas presentes nele.

Na Tabela 1, é possível verificar as 9 empresas selecionadas e uma descrição breve de sua atuação no mercado.

Tabela 1 – Lista das empresas selecionadas para o estudo.

Empresa	Código	Destaque e/ou perfil corporativo
Petrobras	PETR3 / PETR4	Uma das maiores empresas do Brasil e uma das maiores produtoras de petróleo e gás natural do mundo. É líder mundial no desenvolvimento de bacias offshore.
Gerdau	GGBR3 / GGBR4	Maior empresa nacional na produção de aço e uma das principais fornecedoras de aços longos das Américas e aços longos especiais do mundo. Produz também aços planos e minério de ferro.
Vale	VALE3	Uma das maiores mineradoras do mundo. Presente em 30 países, é a maior produtora mundial de minério de ferro, pelotas e níquel.
Companhia Siderúrgica	CSNA3	Empresa multinacional com negócios de siderurgia, mineração, cimento, logística

Nacional (CSN)		e energia. Atua em toda a cadeia produtiva do aço.
Suzano	SUZB3	Empresa com mais de 90 anos de atuação e uma das maiores produtoras verticalmente integradas de papel e celulose de eucalipto da América Latina.
BRF	BRFS3	Empresa com mais de 85 anos, 100 mil funcionários e atuação em 130 países. Possui marcas globais como a Sadia e a Qualy.
Minerva Foods	BEEF3	Empresa produtora e comercializadora de carne in natura e seus derivados, bem como exportação de gado vivo. Faz também o processamento de carne bovina, suína e de aves.
JBS	JBSS3	É a maior produtora de proteínas do mundo e a segunda maior empresa global de alimentos. Possui mais de 400 unidades mundo afora e dezenas de marcas reconhecidas.
Boa Safra Sementes	SOJA3	Companhia líder nacional na produção de sementes de soja. Atende 70% dos estados produtores de soja no Brasil.

Fonte: B3 - Dados do departamento de Relações com Investidores das empresas

4.3 Coletar os preços de fechamento das empresas selecionadas

Neste trabalho, a premissa inicial era utilizar os dados de preço de fechamento ajustado das ações, que é o valor final de uma ação em um dia de negociação, modificado para refletir eventos corporativos, como o pagamento de dividendos, desdobramentos ou grupamentos de ações. Esses ajustes são feitos para garantir que o valor da ação esteja correto ao longo do tempo, permitindo comparações mais justas entre os preços das ações em diferentes períodos. O

cálculo é feito dividindo o preço de fechamento original pelo fator de ajuste, que leva em conta esses eventos.

Entretanto, após uma atualização da API gratuita do Yahoo Finance, os dados de preço de fechamento ajustado ficaram indisponíveis, restringindo o uso a apenas ao preço de fechamento. Todos os dados referente à ações serão retirados do Yahoo Finance¹, que possui uma API gratuita, acessível por meio da biblioteca *yfinance*² em Python.

4.4 Construção de modelos LSTM com apenas os dados históricos das ações e coleta de resultados

Em um primeiro momento, foi construído um modelo com apenas os dados da ação que foram coletados na etapa anterior, para iniciar alguns experimentos e começar ter resultados para analisar a melhor forma de executar o algoritmo e entender seus pontos fracos e fortes. Antes de construir o modelo, os dados foram normalizados com o auxílio da biblioteca *sklearn* e foi utilizado o *MinMaxScaler* que deixa os valores entre 0 e 1, facilitando para o modelo *Long Short-Term Memory* (LSTM) realizar o processamento, treinamento e previsão dos dados.

Na segunda etapa, foi desenvolvido apenas o modelo com os dados de preços ajustado da ação da Petrobrás (PETR4.SA). Ele foi desenvolvido tendo uma rede do tipo LSTM como base, e tem como finalidade prever os preços com base em uma janela de tempo pré-definida. O algoritmo funciona da seguinte forma: a LSTM recebe uma entrada e tem como finalidade prever um valor de acordo com a entrada recebida, sendo assim, ela pode receber uma entrada com mais dados ou menos dados, podendo melhorar ou não a predição. Para a construção desse modelo, a base de dados foi dividida em amostragem de teste e treinamento. Neste trabalho, os dados de treinamento foram o preço ajustado de fechamento da Petrobrás de 2005 até 2021 e os dados de teste foram os de 2021 até 2024.

¹ Disponível em: <https://finance.yahoo.com/>

² Disponível em: <https://github.com/ranaroussi/yfinance>

Depois disso, foram selecionadas algumas empresas para realizar previsões e verificar se a teoria proposta se aplica na realidade e tem eficiência e melhora nas previsões geradas pelo algoritmo LSTM

Na Figura 1 pode-se encontrar o gráfico com os valores reais de fechamento ajustado da ação PETR4.SA, separados em treinamento e teste juntamente com as previsões que foram feitas com o algoritmo. Na Figura 1, está sendo exibido o gráfico com os preços reais da ação da Petrobrás e as previsões do treinamento (vermelho) e do teste (preto). Para verificar a melhor janela, o algoritmo foi executado com janelas desde 1 até 20 dias. A que obteve melhor resultado foi a de 9 dias, que está sendo exibida na figura abaixo.

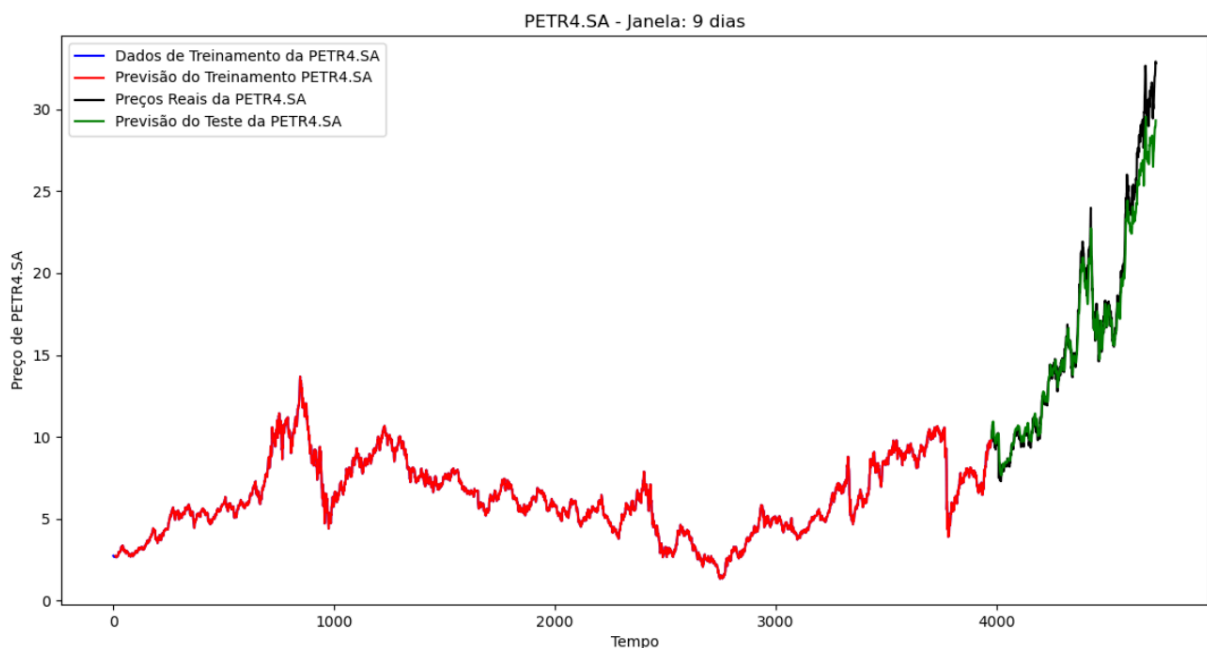


Figura 1: Gráfico da previsão das ações da Petrobrás pelo modelo LSTM com os preços ajustados de fechamento da ação. Fonte: o próprio autor.

Além disso, como dito anteriormente, o algoritmo recebe como entrada uma janela com os valores da ação coletados por um determinado número de dias para realizar a previsão. Como a quantidade de dias utilizados como entrada no modelo LSTM pode influenciar nos resultados, foram feitos experimentos considerando

janelas de 1 a 20 dias com a finalidade de encontrar a melhor janela de tempo para a predição. Além disso, foi observado o erro médio absoluto, erro quadrático médio, raiz quadrada do erro quadrático médio e a métrica R^2 .

Com base nessas duas etapas iniciais, foi possível criar uma base teórica e prática para dar continuidade aos estudos e assim conseguir obter resultados expressivos e aplicáveis ao cotidiano.

Além disso, o trabalho de Silva (2022) colaborou expressivamente para essa pesquisa devido a similaridade dos objetivos e métodos utilizados. O autor realizou uma boa revisão bibliográfica das LSTMs e partiu do pressuposto de que ao utilizar indicadores fundamentais, macroeconômicos e técnicos, o modelo iria absorver o funcionamento do mercado e replicar com melhores resultados e predições e com isso superar o modelo *Naive*. Apesar de ter sido uma base interessante para se espelhar na utilização do *TimeSeriesSplit*, atenção em relação a tunagem dos hiperparâmetros, escolha do algoritmo *Random Search* para tunagem do modelo e pressuposto inicial da adição dos hiperparâmetros, este trabalho possui algumas diferenças que podem ser mencionadas.

Nesta pesquisa, o método de escolha de janelas foi feito de uma maneira mais minuciosa e específica devido ao fato de lidar com 9 empresas diferentes, apesar de serem do mesmo setor de *commodities*. Outra diferença importante foi a utilização do *Keras Tuner* para a otimização e tunagem dos hiperparâmetros dos modelos LSTM's construídos. O Keras Tuner implementa algoritmos de busca sofisticados (como Bayesian Optimization, Hyperband, Random Search) que exploram o espaço de hiperparâmetros de forma mais metódica e abrangente do que seria viável manualmente. Isso aumenta significativamente a probabilidade de encontrar configurações de alto desempenho. Por fim, no trabalho de Silva, foi escolhido trabalhar com S&P500 que sofre influência de outros fatores e possui uma realidade um pouco distante do setor das *commodities* no Brasil.

Erro médio absoluto (MAE)

O erro médio absoluto é uma importante medida de erro que verifica a qualidade das previsões. Ele mede a média da diferença absoluta entre os valores previstos pelo modelo e os valores observados:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

- n é o número de amostras
- y é o valor observado para cada amostra
- p é o valor previsto pelo modelo para cada amostra
- $| |$ representa o valor absoluto
- i representa cada observação

Erro quadrático médio (MSE)

O erro quadrático médio é semelhante ao erro médio absoluto mas ele é sensível a erros maiores, sendo importante para detectar valores discrepantes que não se encaixam no modelo:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- n é o número de amostras
- y é o valor observado para cada amostra
- \hat{y} representa o valor previsto para cada amostra
- i representa cada observação

Raiz quadrada do erro quadrático médio (RMSE)

A raiz quadrada do erro médio é apenas a raiz quadrada do MSE, onde o erro retorna à unidade de medida do modelo (no MSE, a unidade de medida é quadrática). É frequentemente usada em séries temporais por ser mais sensível a erros maiores.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

- n é o número de amostras

- y é o valor observado para cada amostra
- \hat{y} representa o valor previsto para cada amostra
- i representa cada observação

R-Quadrado

O R^2 (coeficiente de determinação) é uma medida estatística que indica o quão bem os valores observados em uma amostra são explicados por um modelo de regressão. Ele avalia a proporção da variância total dos dados que é explicada pelo modelo ajustado.

Dito isso, a Figura 2 permite visualizar com maior facilidade como a LSTM funciona para cada janela de tempo dentro do intervalo selecionado (1 até 20). Ou seja, os modelos são treinados 20 vezes, uma vez com cada janela de tempo na entrada e são salvos seus valores por janela de tempo junto com as medidas de erro. Como é nítido na Figura 2, a janela de 9 dias levou ao melhor desempenho, considerando as medidas de erro e seus resultados, visando o que apresenta menores medidas de erro e maior valor de precisão, o que garante maior confiabilidade nessa janela.

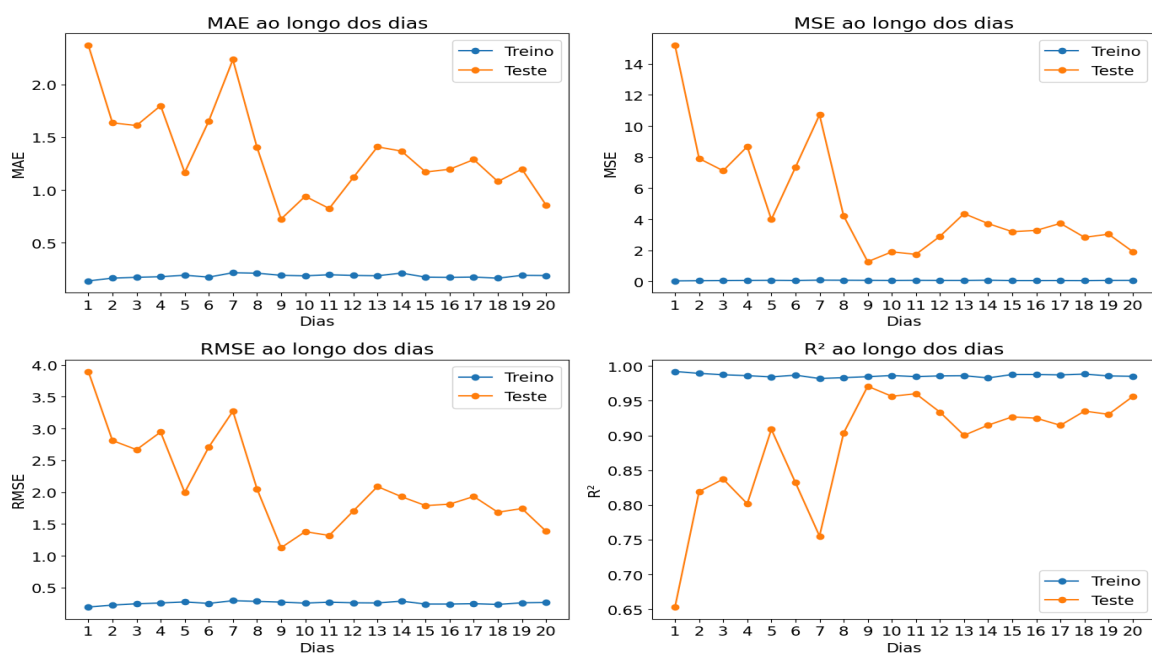


Figura 2: Resultados das métricas de erro obtidas com cada janela de tempo utilizada como entrada para a LSTM. Fonte: o próprio autor.

Entretanto, somente um experimento com esse intervalo de janelas não seria suficiente para garantir uma boa escolha. Com isso, foi repetido o experimento descrito acima, porém utilizando validação cruzada (k fold) para analisar novamente qual é o melhor intervalo de tempo para alimentar o modelo LSTM. O processo começa dividindo o conjunto de dados original em k subconjuntos (ou "*folds*") de tamanho aproximadamente igual. O valor de k é escolhido pelo usuário, sendo comum usar valores como 5 ou 10.

Uma vez que os dados são divididos, o processo de treinamento e validação é repetido k vezes. Em cada iteração, um dos k subconjuntos é usado como conjunto de validação, enquanto os $k - 1$ subconjuntos restantes são usados como conjunto de treinamento. Por exemplo, se $k = 5$, os dados são divididos em 5 *folds*. Na primeira iteração, o primeiro *fold* é usado para validação e os outros 4 *folds* são usados para treinamento. Na segunda iteração, o segundo *fold* é usado para validação e os outros 4 *folds* são usados para treinamento, e assim por diante, até que cada *fold* tenha sido usado uma vez como conjunto de validação.

Em resumo, a validação cruzada *K-Fold* é uma técnica poderosa para avaliar a capacidade de generalização de um modelo de aprendizado de máquina. Ela garante que o modelo seja testado em diferentes partes dos dados, proporcionando uma avaliação mais robusta e confiável do seu desempenho.

A Figura 3 apresenta os resultados de cada *fold*, onde é possível observar as métricas de erro calculadas e analisar o desempenho do modelo ao longo das janelas de dias definidos.

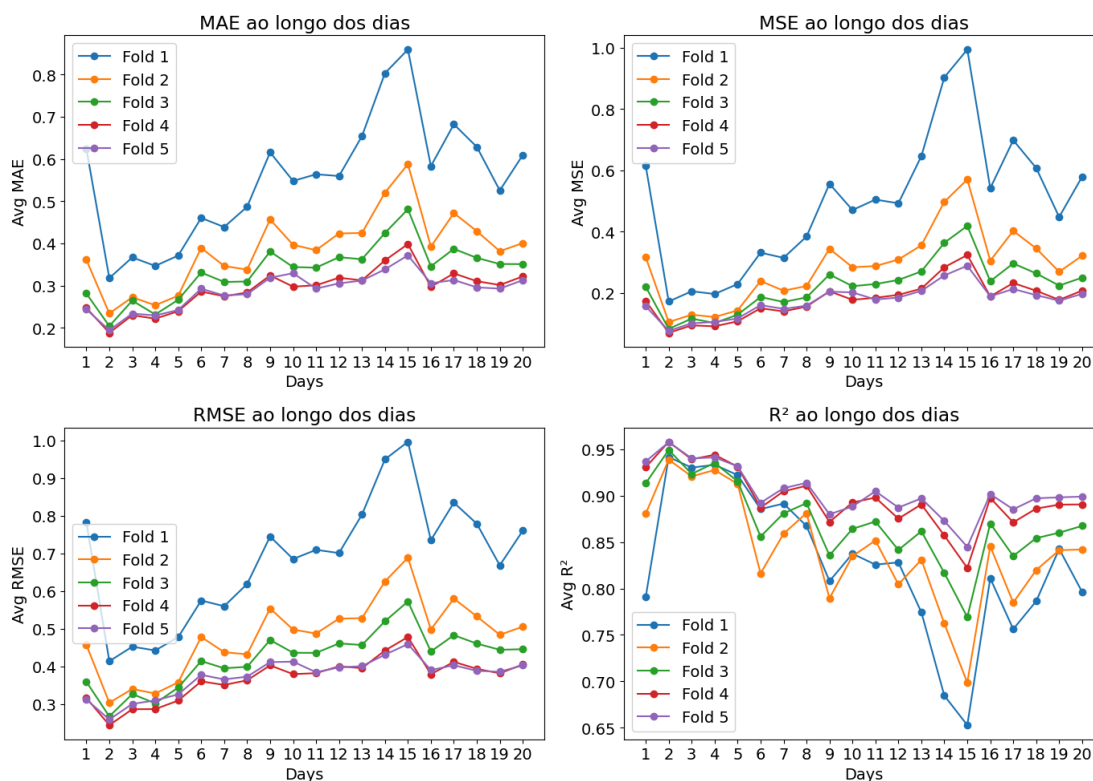


Figura 3: Resultados das métricas de erro para cada partição utilizando validação cruzada (k fold). Fonte: o próprio autor.

Na Figura 3 é possível observar que o melhor desempenho foi na janela de 2 dias e com o fold 5. Ao analisar o gráfico, é possível concluir que o modelo está se auto ajustando conforme tem mais dados disponíveis à medida que os *folds* aumentam, comprovando uma boa generalização dos dados. Com isso, ocorreu uma diferença significativa entre os resultados. Dessa forma, considerando que a validação cruzada K -Fold é uma técnica mais robusta, é mais seguro utilizar o resultado obtido da mesma, já que o primeiro experimento pode ter apresentado algum viés quando os dados foram separados em treinamento e teste.

Entretanto, esse resultado estava vinculado somente aos dados da ação Petrobrás, sendo necessário realizar análises sobre os outros ativos para verificar se esse padrão se repetia. Ao repetir esse procedimento para os demais ativos, obteve-se esse resultado:

Comparison of Best Results Across Tickers									
	Ticker	MAE	MAE_Days	RMSE	RMSE_Days	MSE	MSE_Days	R ²	R ² _Days
0	PETR4.SA	0.3059	2	0.4337	2	0.1881	2	0.9956	3
1	VALE3.SA	0.9276	1	1.3143	1	1.7274	1	0.9918	4
2	BRFS3.SA	0.4017	1	0.5781	1	0.3342	1	0.9918	1
3	JBSS3.SA	0.3466	3	0.5131	3	0.2632	3	0.9894	3
4	SOJA3.SA	0.2105	1	0.3068	1	0.0941	1	0.9643	1
5	SUZB3.SA	0.8100	3	1.0804	3	1.1674	3	0.9781	3
6	GGBR4.SA	0.2902	1	0.4086	1	0.1669	1	0.9893	1
7	CSNA3.SA	0.3946	1	0.5540	1	0.3069	1	0.9938	1

Figura 4: Resultados das melhores janelas para cada ativo. Fonte: o próprio autor

Sendo assim, foi definido um intervalo entre 1 e 4 dias para os experimentos seguintes, garantindo assim que abrangeria os potenciais melhores resultados.

4.5 Estudo do setor empresarial

Esta etapa é voltada mais para literatura e análise do setor empresarial com a finalidade de compreender melhor quais são os indicadores econômicos que influenciam o desempenho operacional das empresas presentes nele e buscar selecionar os melhores indicadores. Na literatura, frequentemente é mencionada a “a maldição dos recursos naturais” (Veríssimo, Xavier, 2014) que parte do pressuposto que países ricos em recursos naturais apresentam taxas menores de crescimento econômico, como Venezuela, Rússia e Nigéria, quando comparados a países pobres em recursos naturais, como Hong Kong, Coreia, Japão e Cingapura. A maldição dos recursos naturais está vinculada ao argumento de que o súbito aumento da renda decorrente da exploração daqueles recursos (derivado de um aumento dos preços destes produtos ou da descoberta de novas fontes) cria uma falsa ideia de segurança e enfraquece a necessidade percebida de investimento e de promoção de estratégias de crescimento. Nestes termos, países ricos em recursos naturais tendem a desenvolver políticas protecionistas, com burocracia e

ineficiência na utilização dos recursos (fraqueza institucional), o que acarreta menores taxas de investimento, e, conseqüentemente, menor crescimento econômico (Veríssimo, Xavier, 2014).

Portanto, a hipótese a ser testada é que algumas *commodities* ao serem importadas, geram diferentes efeitos na taxa de câmbio e no crescimento econômico. Como o Brasil é um país que centraliza investimentos no setor das *commodities* e as exportações refletem muito no seu desempenho econômico, é de suma importância desenvolvimento de ferramentas que auxiliem a prever momentos de baixa para definir a melhor estratégia de investimento.

Em seus resultados, Veríssimo & Xavier (2014) conseguiram comprovar que diferentes *commodities* geram diferentes efeitos na taxa de câmbio e na taxa de crescimento econômico, coletando evidências fracas dessa relação de quanto mais recursos naturais, menor será o crescimento desse país, sendo excelentes indicadores para serem utilizados neste trabalho.

4.6 Seleção de indicadores macro/microeconômicos com possível relevância

Após a etapa anterior, será possível selecionar os principais indicadores que têm maior probabilidade de influenciar no dia a dia das empresas e que possam ajudar na melhora da predição do algoritmo.

Em sua tese, Timm (2023) buscou compreender quais indicadores influenciam as empresas de mineração e siderurgia, ligadas às *commodities*, e analisar os resultados para obter conclusões. Ele observou indicadores como P/L, ML, ROE, VPA, P/VPA,.

$$P/L = \frac{\text{Preço}}{\text{Lucro}}$$

$$ML = \left(\frac{\text{Lucros Líquidos}}{\text{Vendas Líquidas}} \right) \times 100$$

$$ROE = \frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}}$$

$$VPA = \frac{\text{Patrimônio Líquido}}{\text{Número de Ações}}$$

$$P/VPA = \frac{\text{Preço da Ação}}{VPA}$$

Em seu trabalho, ele conseguiu comparar dois setores de *commodities* diferentes e como os indicadores afetam os setores. Em sua conclusão, foi verificado que o setor de mineração teve números superiores em relação ao de siderurgia, tanto nos indicadores quanto na análise fundamentalista. Portanto, como neste trabalho está sendo estudado o setor das *commodities*, ele abrange alguns outros sub setores, como agrícola, mineração e energia, dentre outros, por isso é importante adotar a metodologia do autor para conseguir filtrar os indicadores que mais afetam as empresas e obter melhores resultados.

Em sua conclusão, o autor afirma que o estudo dele se tornou não apenas um comparativo entre empresas de mineração e siderurgia, mas sim uma fonte valiosa de informação para investidores e stakeholders. Além disso, o uso de indicadores financeiros e fundamentalistas fornecem informações cruciais para a tomada de decisões em investimentos. Estratégias de investimentos baseadas a curto prazo podem se beneficiar bastante do uso dessas informações.

4.7 Coleta de dados dos indicadores

Para a coleta de dados dos indicadores, será crucial o uso de APIs gratuitas existentes em Python ou até mesmo técnicas de *web scraping*. Por exemplo, a taxa de juros poderia ser obtida por meio de uma requisição no site da Receita Federal por meio da biblioteca *requests*³ em Python. Sendo assim, para os outros

³ Disponível em: <https://pypi.org/project/requests/>

indicadores econômicos é importante pesquisar previamente como extrair essas informações da internet da melhor maneira possível.

4.8 Verificação de relevância dos indicadores

Após a extração dos dados, será preciso verificar a relevância dos indicadores e se eles possuem alguma relação com a série temporal em questão. Para investigar a relação linear entre os indicadores micro e macroeconômicos selecionados para cada ativo, utilizou-se o coeficiente de correlação de Pearson. Esta métrica padrão permite quantificar a força e a direção da associação entre pares de variáveis, sendo um passo fundamental para entender possíveis movimentos e identificar redundâncias nos dados antes de etapas subsequentes de modelagem ou análise.

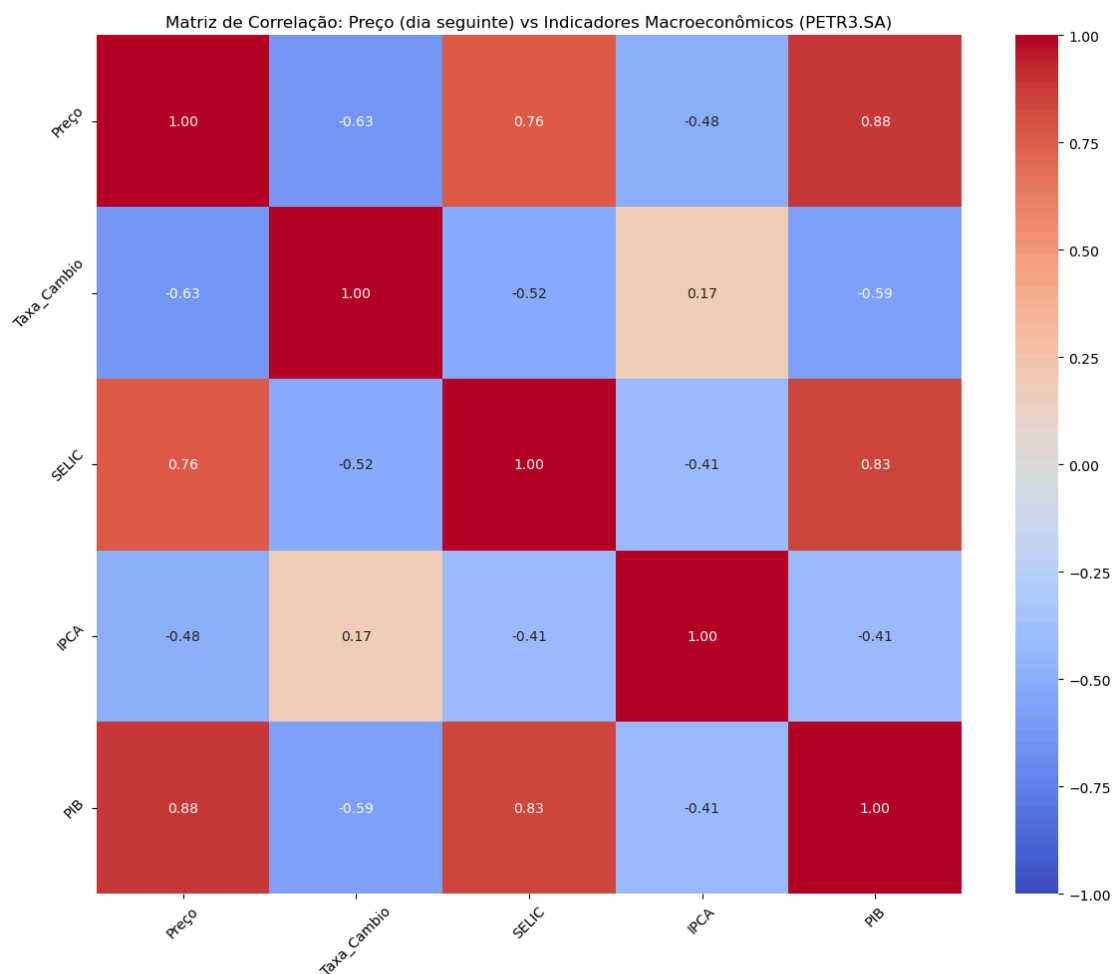


Figura 5: Matriz de Correlação dos preços da PETR3.SA com os indicadores macroeconômicos. Fonte: o próprio autor

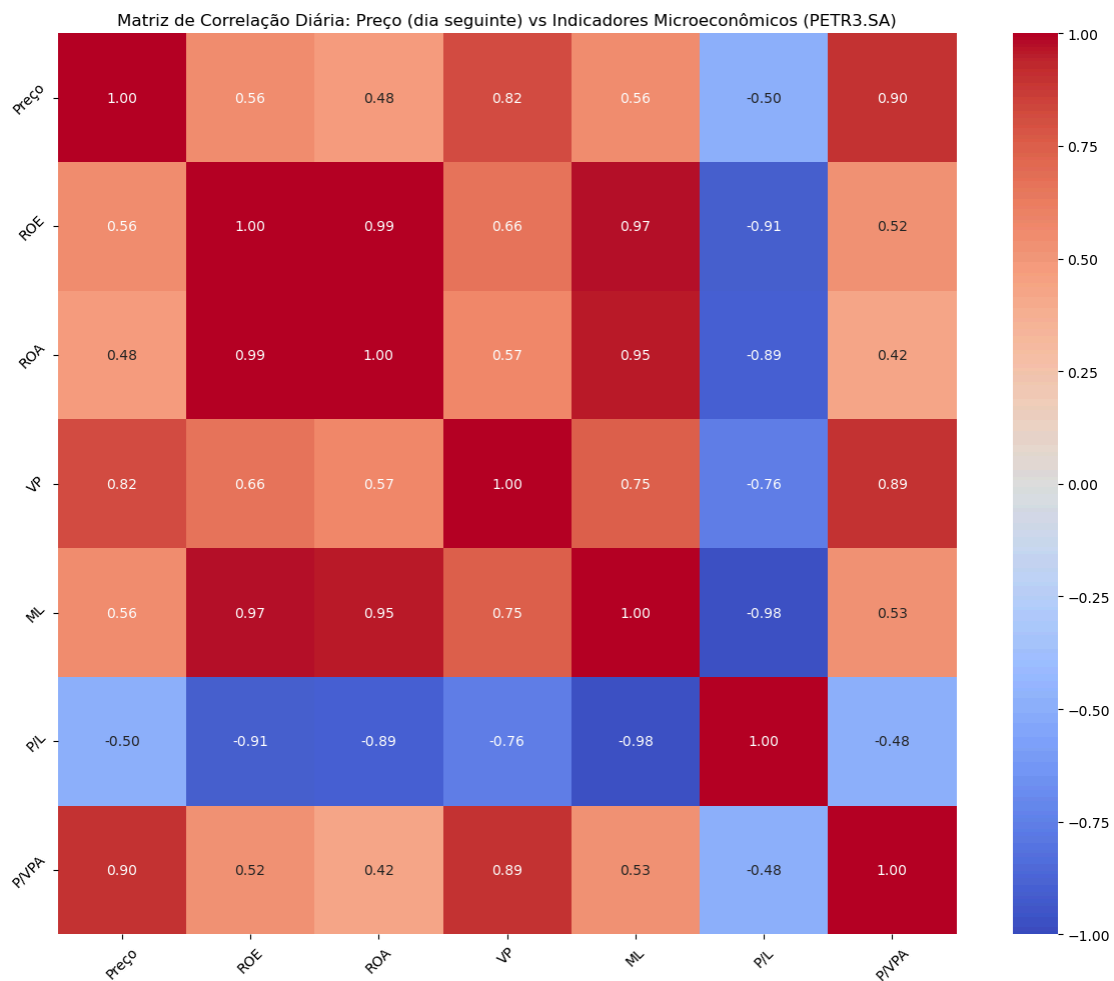


Figura 6: Matriz de Correlação dos preços da PETR3.SA com os indicadores microeconômicos.
Fonte: o próprio autor

Para construir essas matrizes, foi criado um *dataframe* que relacionava os preços de fechamento de um dia com os valores dos indicadores do dia anterior, refletindo diretamente como funciona o mercado atual. Como foi observado, não existe um padrão entre os indicadores, reforçando a ideia de que por mais que as empresas sejam do mesmo setor econômico, elas possuem realidades diferentes e não é possível separar os mais relevantes de forma que agregue as 9 empresas. O restante das matrizes de correlação estão disponíveis no apêndice para maiores detalhes.

4.9 Construção de modelos LSTM com dados das ações combinados aos indicadores e coleta de resultados

Nessa etapa, será construído um modelo otimizado tanto na escolha da melhor janela quanto nos hiperparâmetros já ajustados de acordo com os resultados experimentais obtidos com o modelo anterior, porém serão adicionados os indicadores previamente selecionados e pesquisados para verificar sua influência na prática e buscar a melhor compreensão daquele setor empresarial.

4.10 Desenvolvimento de uma estratégia de investimentos baseada nos modelos criados

Parou aqui

4.11 Avaliação da estratégia desenvolvida com simulações de investimento

Elaborada a estratégia, ela deverá ser testada na prática para verificar sua eficácia e aplicabilidade no cotidiano. Para isso, será estruturada uma simulação com os valores do ativo e as previsões geradas pelo modelo.

4.12 Análise de resultados

Esta será a última etapa do trabalho e nela será verificado tudo o que foi proposto até então. Será possível verificar se a adição dos indicadores econômicos melhorou ou não a predição, qual a efetiva aplicabilidade de um modelo como os desenvolvidos no dia a dia, por meio de uma estratégia de investimentos, e serão feitas comparações com experimentos de trabalho correlatos existentes na literatura.

5. Resultados

6. Conclusão

7. Referências Bibliográficas

- ASSAF NETO, Alexandre. **Finanças corporativas e valor**. 7. ed. São Paulo: Atlas, 2014
- ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS (ANBIMA). **Desempenho do ano. Relatório Anual 2023**. Disponível em: <https://www.anbima.com.br/relatorioanual/2023/desempenho-do-ano.html>.
- B3 – Brasil, Bolsa, Balcão. **Dados do departamento de Relações com Investidores das empresas**. Disponível em: <https://www.b3.com.br>.
- BARROS, T. S. **Análise Técnica e Fundamentalista: Ensaio Sobre os Métodos de Análise**. Revista de Administração e Negócios da Amazônia, Universidade Federal de Ouro Preto, Minas Gerais, v. 7, n. 2, p. 39-63, maio, 2015. Disponível em: https://www.repositorio.ufop.br/bitstream/123456789/6872/1/ARTIGO_An%C3%A1liseT%C3%A9cnicaFundamentalista.pdf. Acesso em: 17 jan. 2024.
- BOMFIM, Paulo Roberto Clemente Marques; MACEDO, Marcelo Álvaro da Silva; MARQUES, José Augusto Veiga da Costa. **Indicadores financeiros e operacionais para a avaliação de desempenho de empresas do setor de petróleo e gás**. Revista de Contabilidade e Organizações, v. 16, n. 1, 2019.
- BRANDÃO, Juliana Carvalho; VOGT, Camila Moura. **Os efeitos macroeconômicos do superciclo de commodities e a influência da China na economia brasileira**. Revista Tempo do Mundo, Brasília, v. 24, n. 1, p. 284-313, dez. 2020. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/13395/1/Tempo_Mundo_n24_Artigo10_efeitos_macroeconomicos.pdf. Acesso em: 2 dez. 2024.
- COMISSÃO DE VALORES MOBILIÁRIOS (CVM). Instrução CVM nº 409, de 18 de agosto de 2004. Consolidação das normas sobre a constituição, administração e funcionamento dos fundos de investimento. Disponível em: <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/instrucoes/anexos/400/inst409consolid.pdf>.
- GERBELLI, Luiz Guilherme. **Quais são os setores e as empresas mais importantes da Bolsa brasileira? Veja em gráficos**. O Estado de S. Paulo, São Paulo, 19 out. 2023. Disponível em: <https://www.estadao.com.br/economia/empresas-setores-peso-bolsa-b3/>. Acesso em: 2 dez. 2024.
- MALKIEL, G. B.; FAMA, E. F. Efficient capital markets: **A review of theory and empirical work**. The Journal of Finance, v. 25, n. 2, p. 383-417.
- MUELLER, A. **Uma Aplicação de Redes Neurais Artificiais na Previsão do Mercado Acionário**. Dissertação (Mestrado em Engenharia) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 1996.
- RIBEIRO, Maitê Garcia Cruz; MACEDO, Marcelo Álvaro da Silva; MARQUES, José Augusto Veiga da Costa. **Análise da relevância de indicadores financeiros e não financeiros na avaliação de desempenho organizacional: um estudo exploratório**

no setor brasileiro de distribuição de energia elétrica. Revista de Contabilidade e Organizações, v. 6, n. 15, p. 60-79, 2012.

SANTOS, G. T. **Análise Fundamentalista e Aplicação de Técnicas Econométricas para a Previsão de Séries Temporais Financeiras.** Tese (Graduação em Estatística) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, 2023.

TIMM, Bernardo Steffen. **Análise comparativa de indicadores financeiros na avaliação de empresas do setor de mineração e siderurgia, listadas na bolsa de valores brasileira.** 2023. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) – Universidade Federal de Santa Maria, Santa Maria. Disponível em: <https://portal.ufsm.br/documentos/publico/autenticacao/assinaturas.html>.

VERÍSSIMO, Michele Polline; XAVIER, Clésio Lourenço. **Tipos de commodities, taxa de câmbio e crescimento econômico: evidências da maldição dos recursos naturais para o Brasil.** Revista de Economia Contemporânea, Rio de Janeiro, v. 18, n. 2, p. 267-295, 2014. Disponível em: <http://dx.doi.org/10.1590/141598481825>.

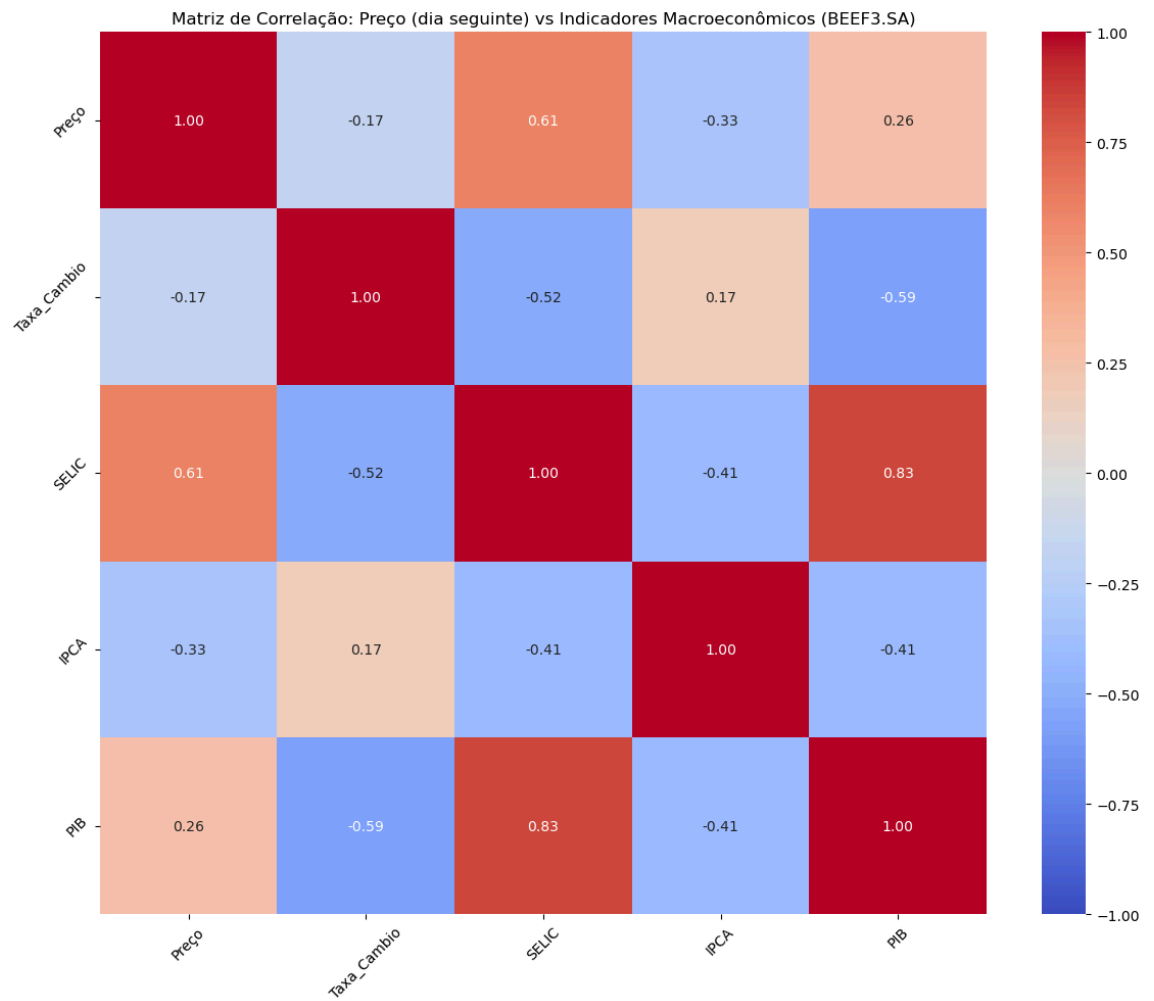
SILVA, André Filipe Ferreira Lourenço da. **Redes Neurais LSTM em Python: Previsão do S&P500.** 2022. Dissertação (Mestrado em Finanças) – ISCTE Business School, Instituto Universitário de Lisboa, Lisboa, 2022. Disponível em: <https://repositorio.ulisboa.pt/handle/10451/61679>

FRAGA, Breno Mendes Cardoso. **Integração de Redes Neurais do Tipo LSTM em Otimização de Portfólios de Investimento.** 2021. Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) - Escola Politécnica da Universidade de São Paulo. Disponível em: <https://repositorio.usp.br/directbitstream/b581998b-bd1d-4c27-b6ea-90099c027c79/BRENO%20MENDES%20CARDOSO%20FRAGA%20PRO2021.pdf>

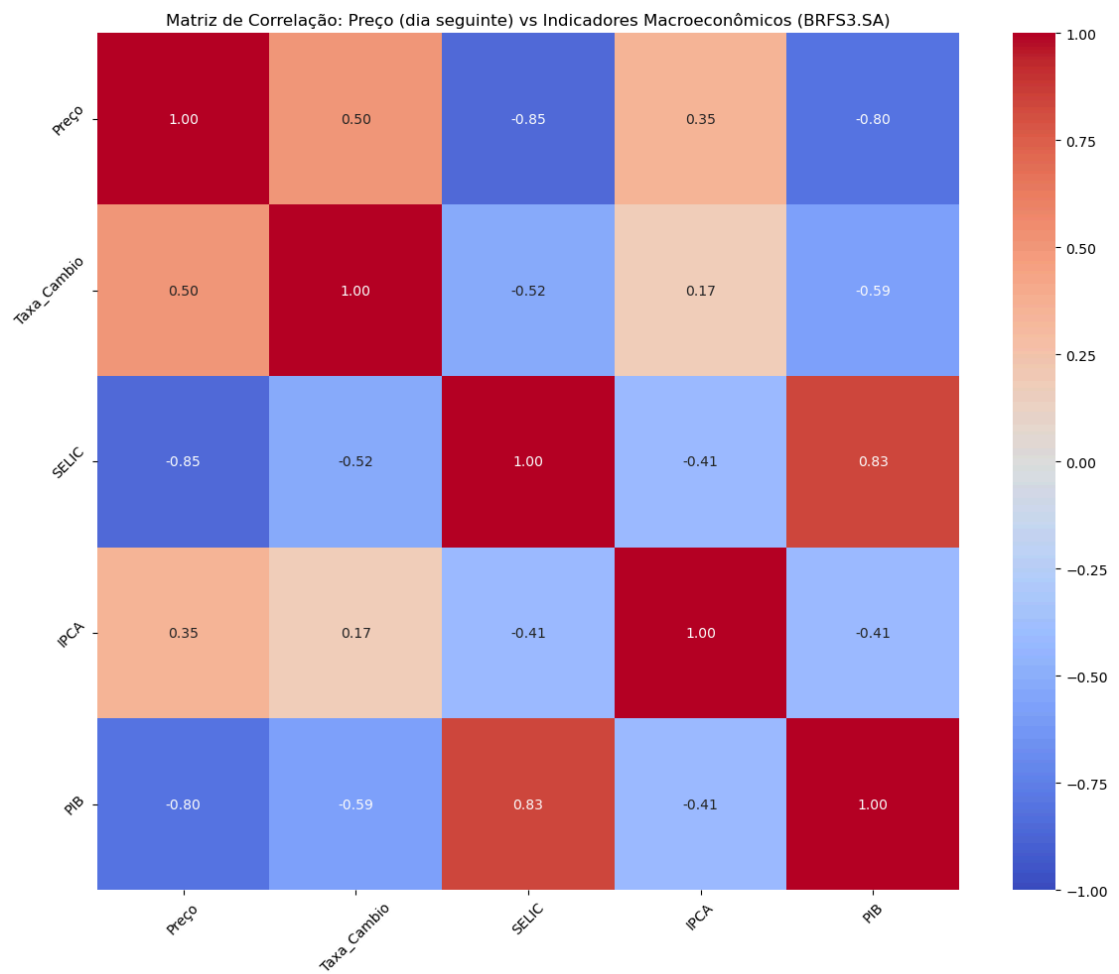
8. Apêndices

8.1 Matrizes de Correlação (Preço x Indicadores Macroeconômicos)

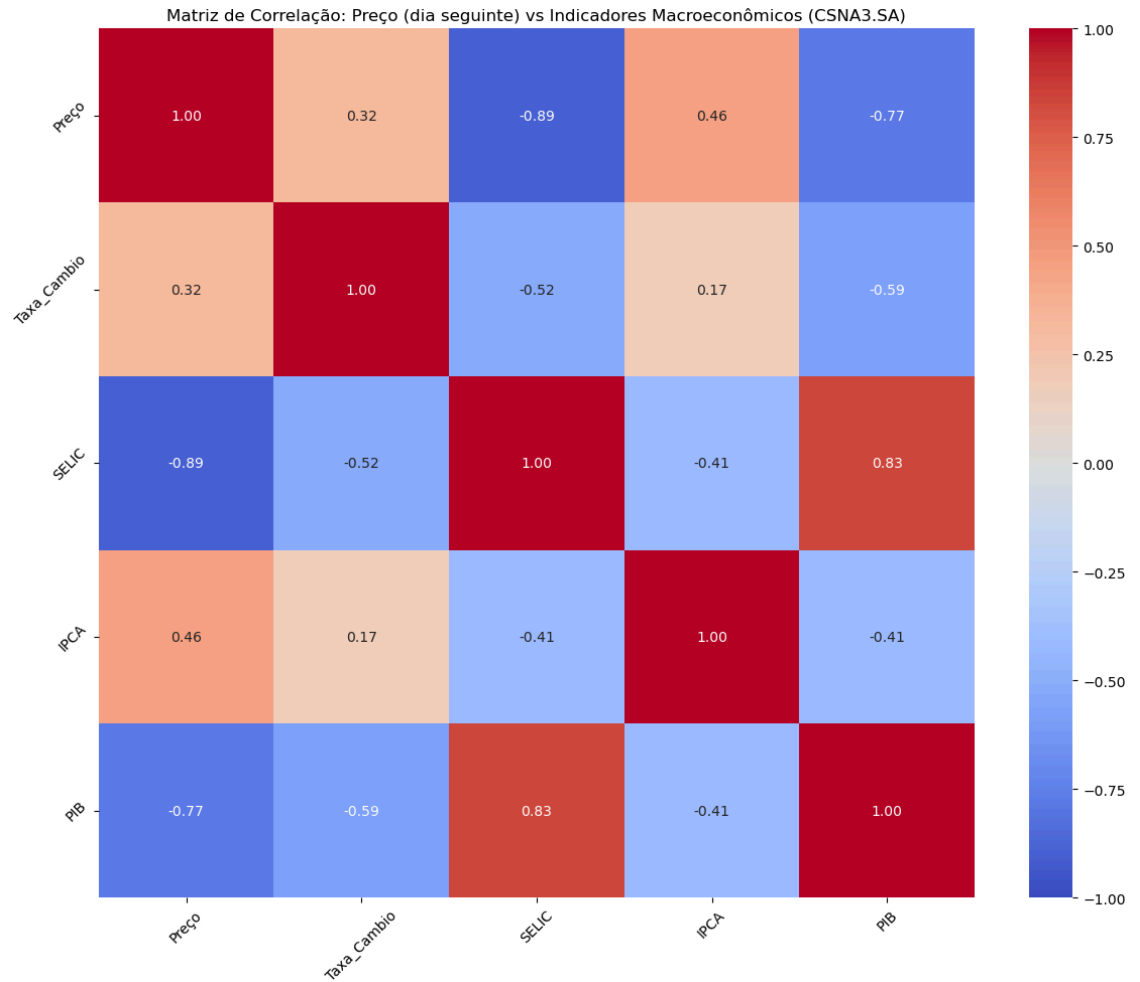
BEEF3.SA



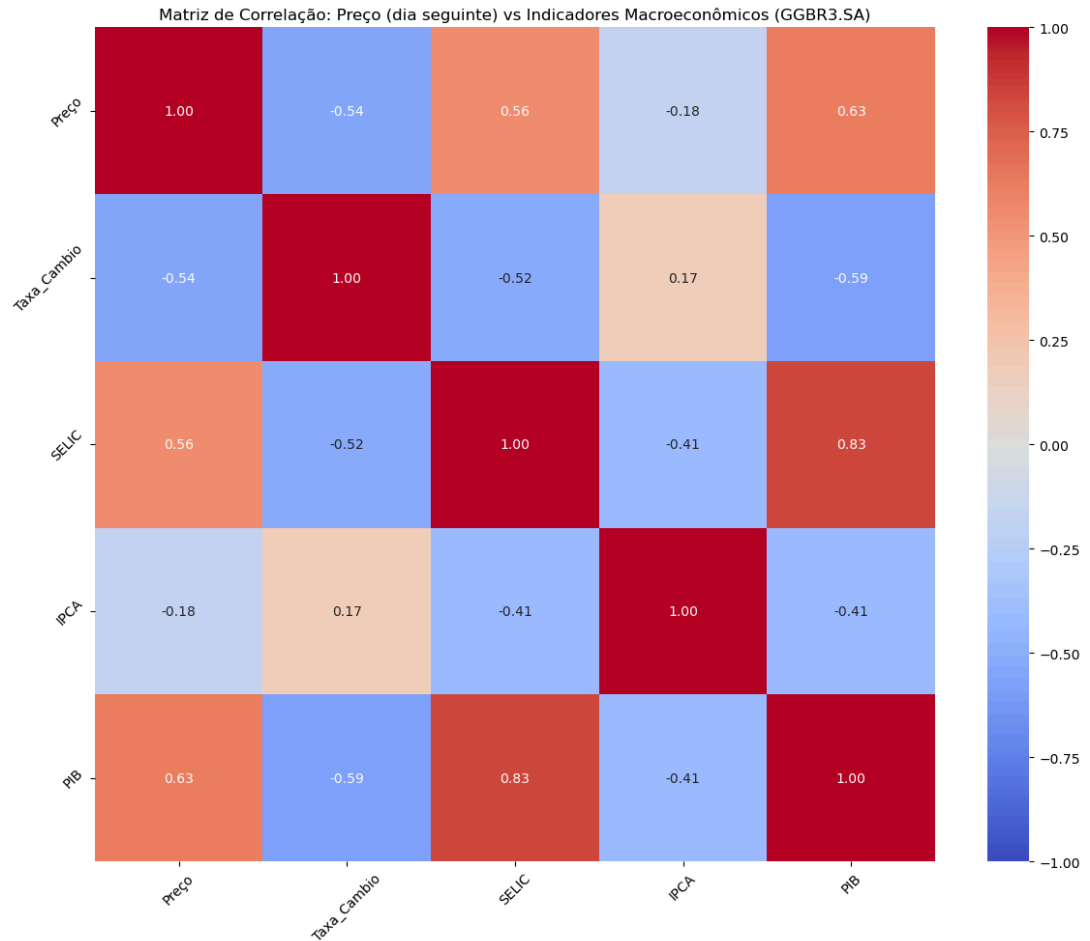
BRFS3.SA



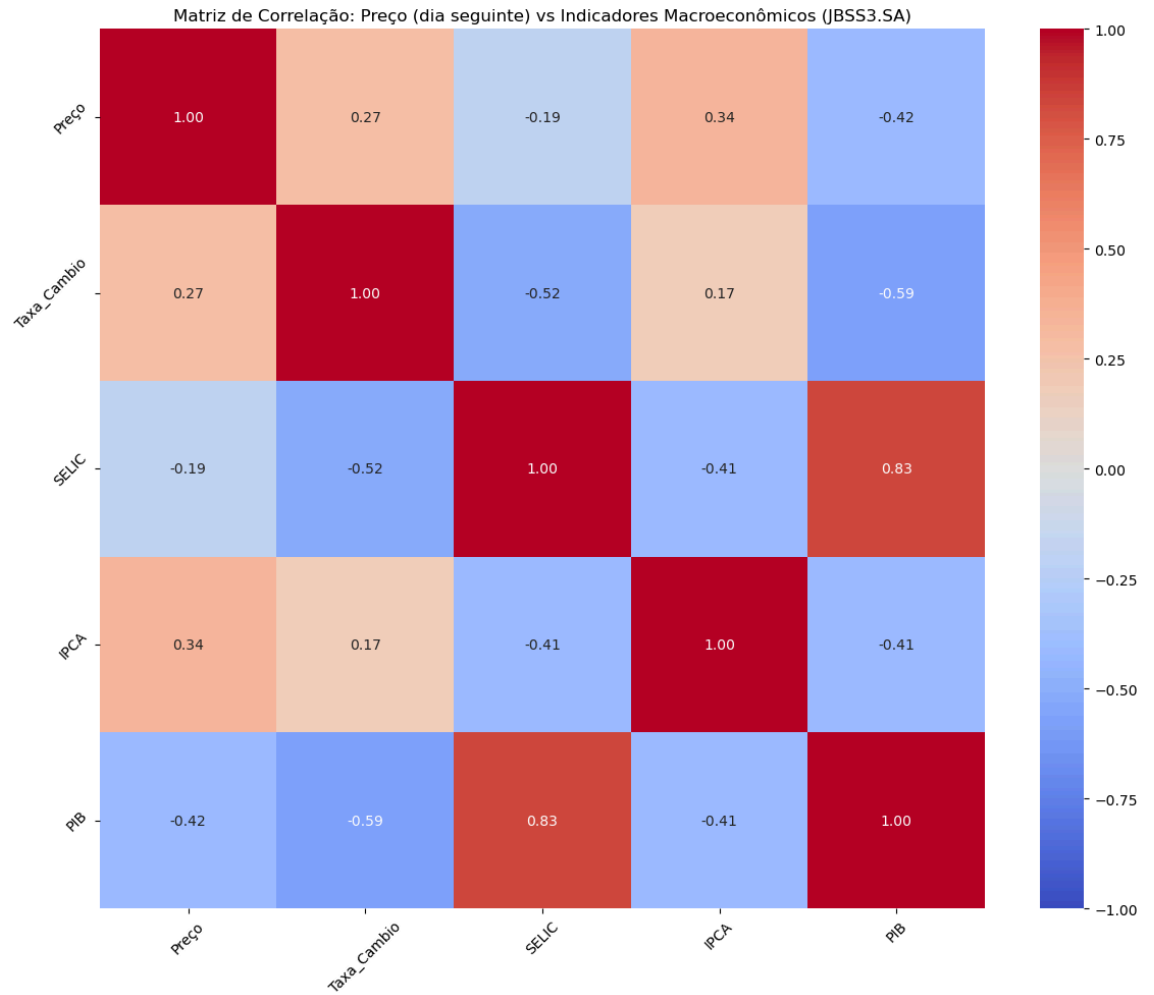
CSNA3.SA



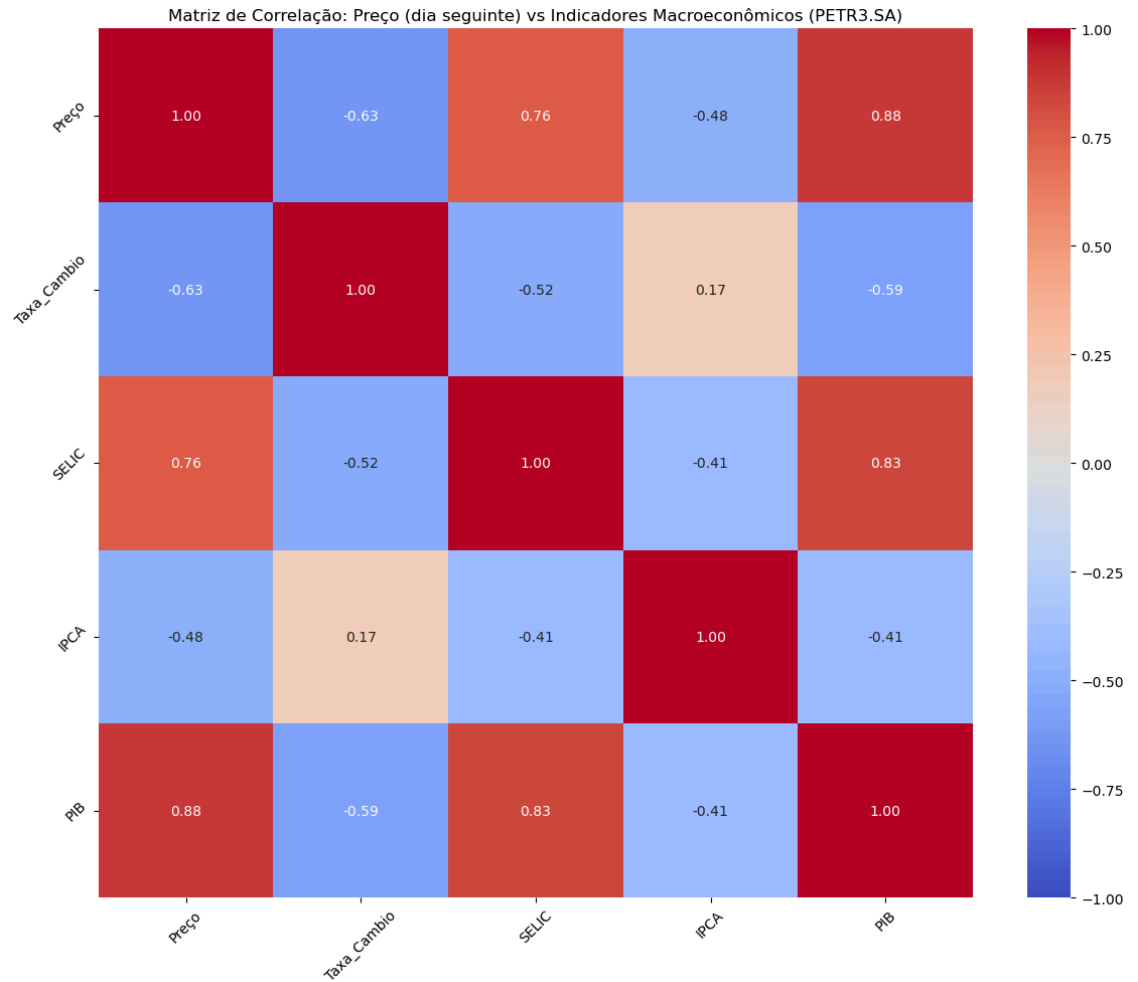
GGBR3.SA



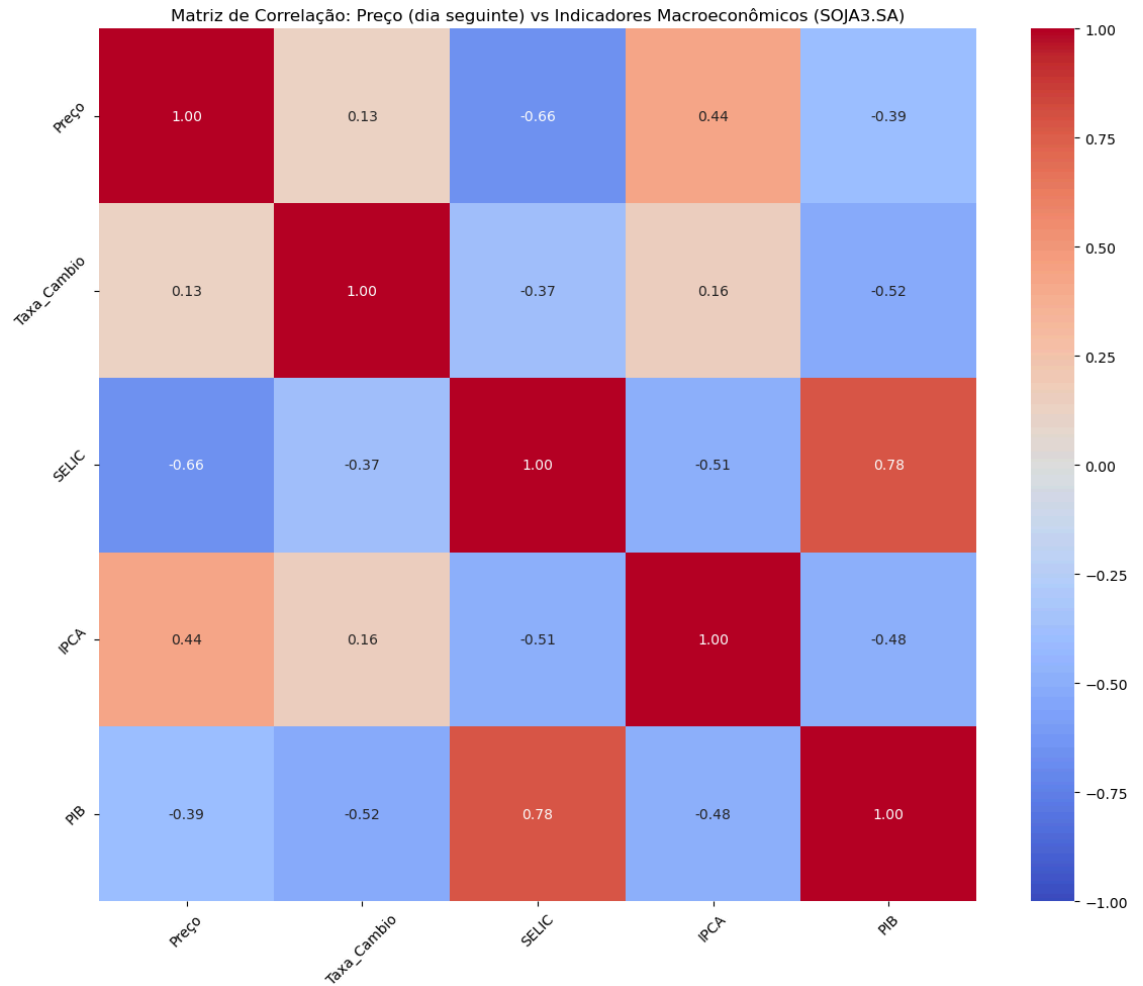
JBSS3.SA



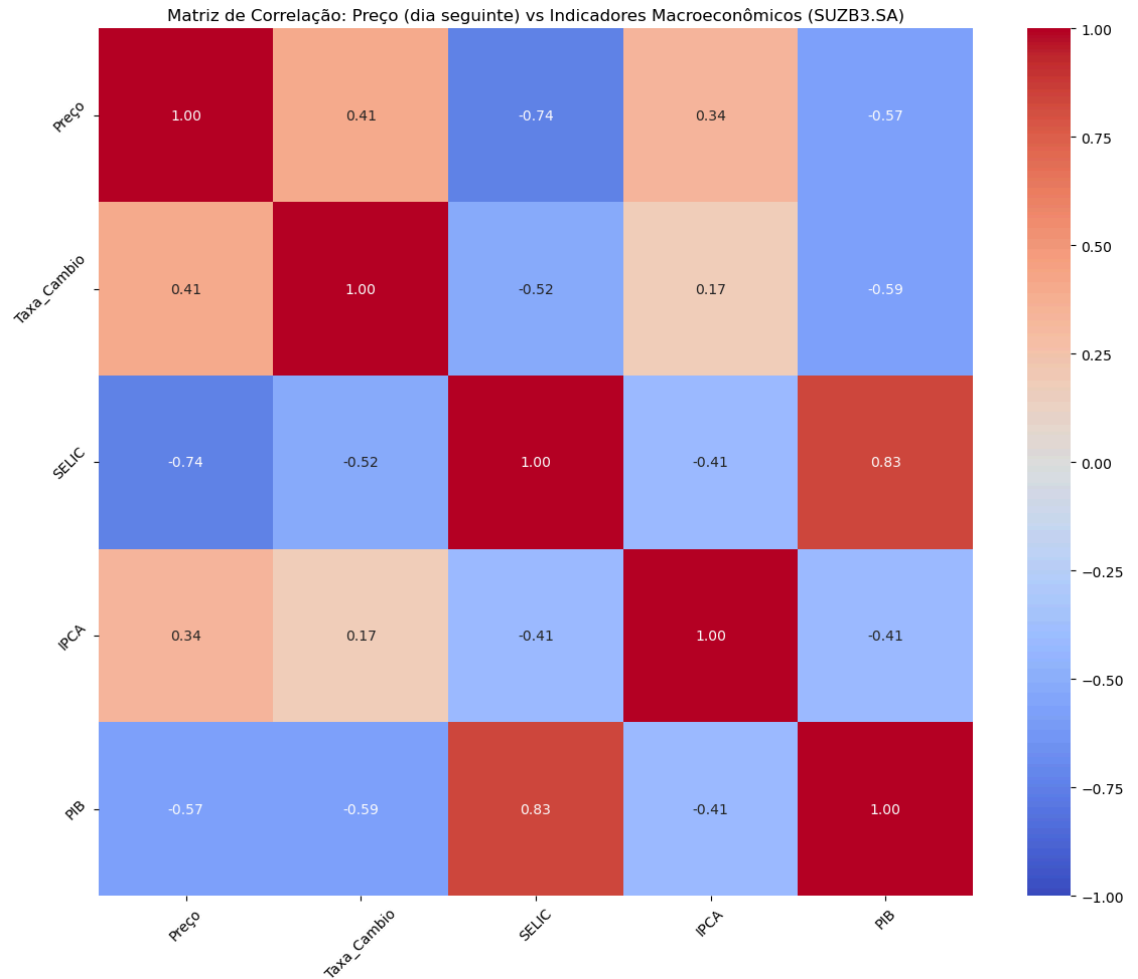
PETR3.SA



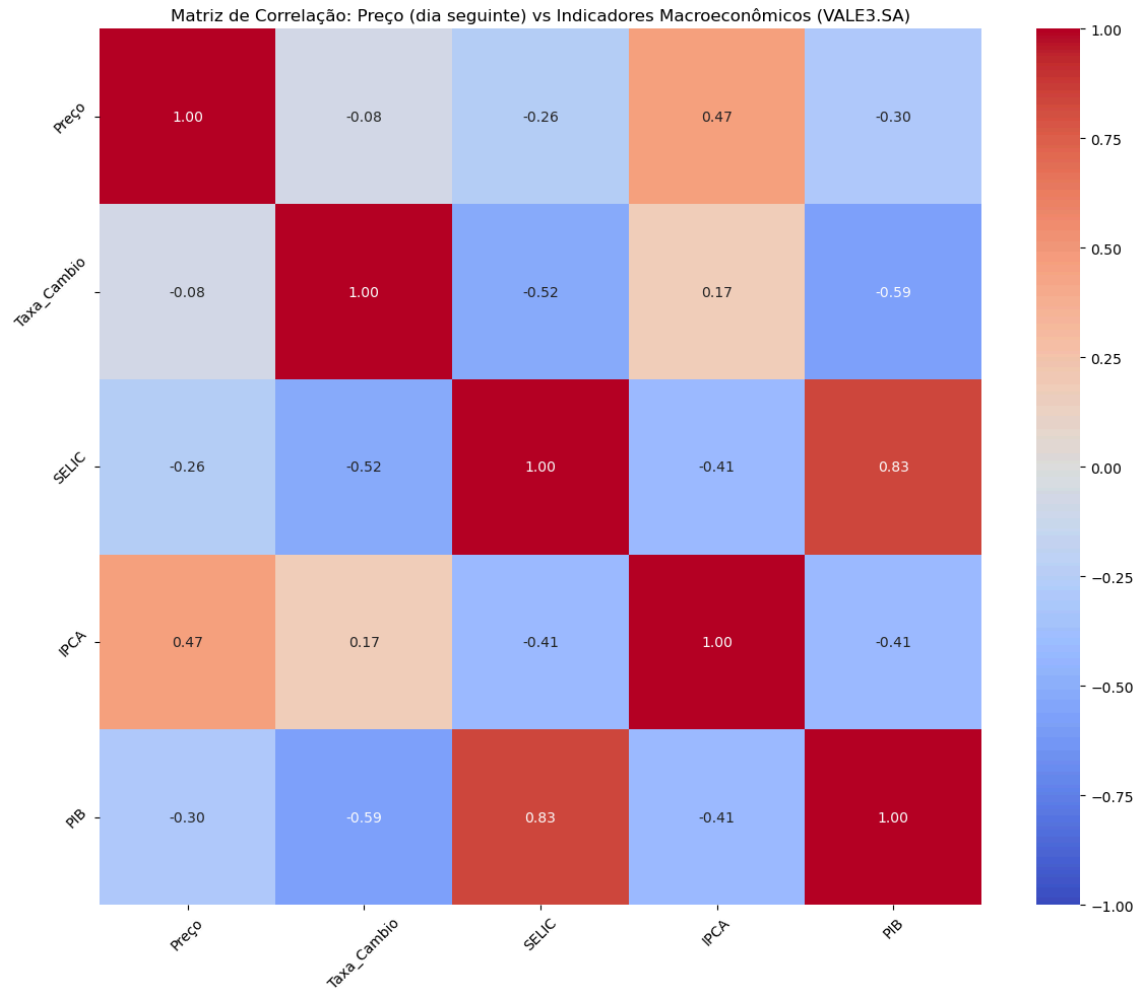
SOJA3.SA



SUZB3.SA

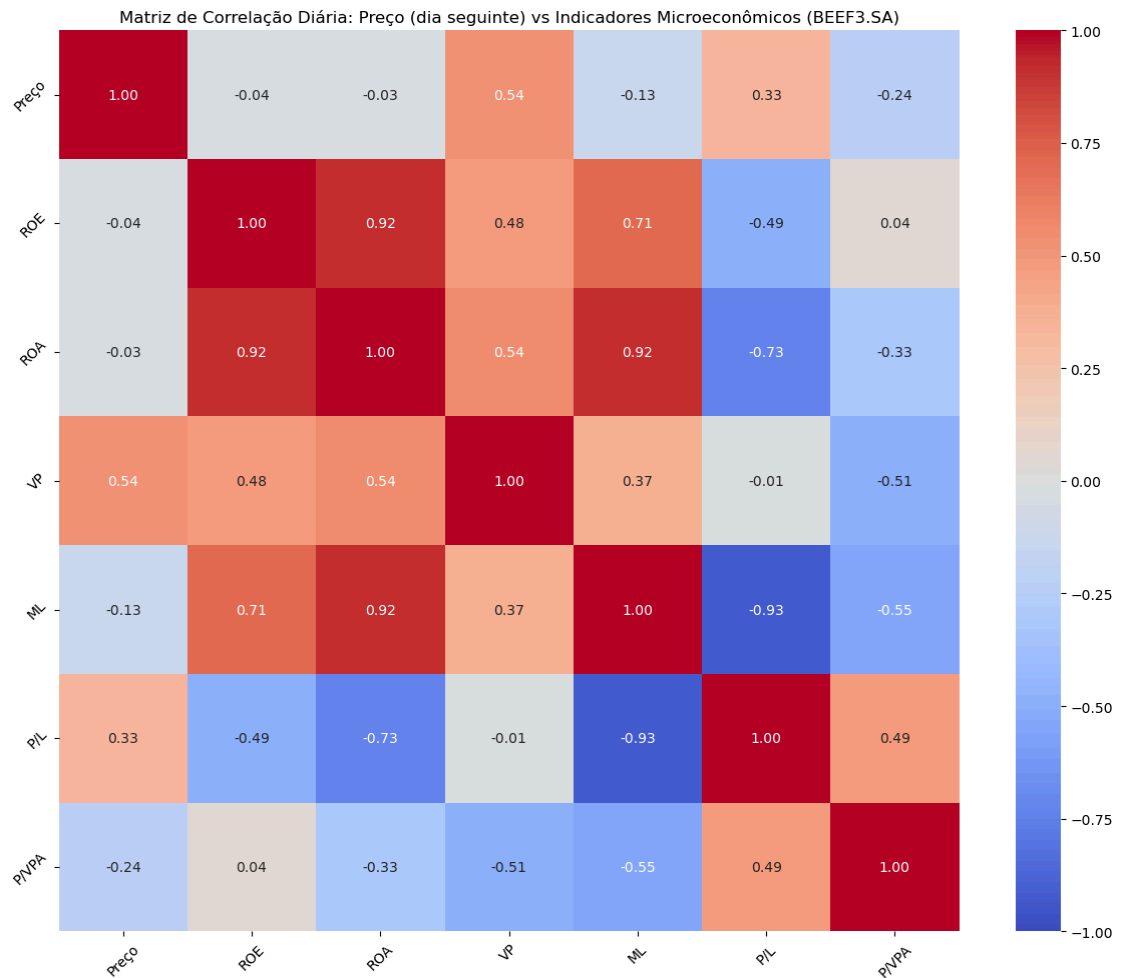


VALE3.SA

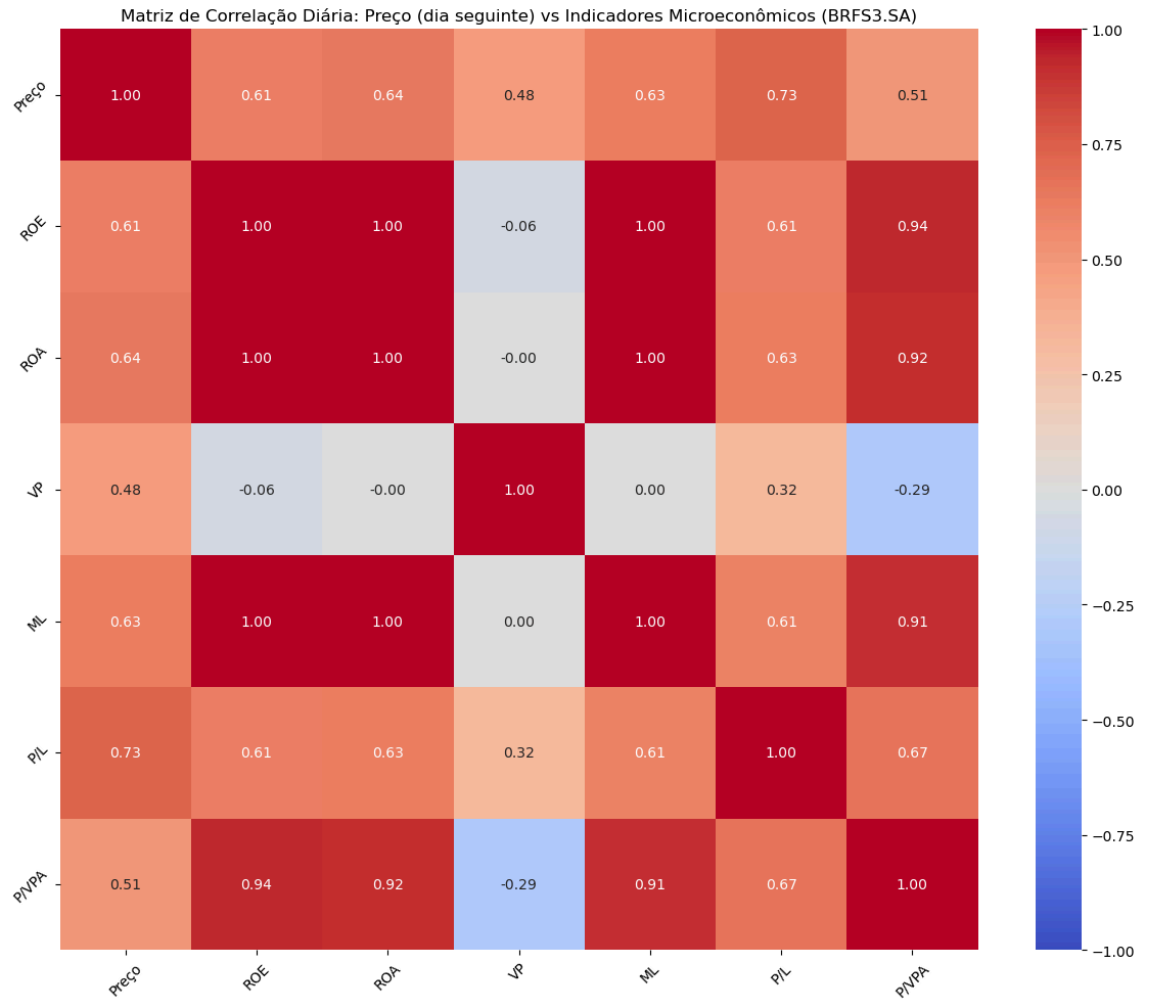


8.2 Matrizes de Correlação (Preço x Indicadores Microeconômicos)

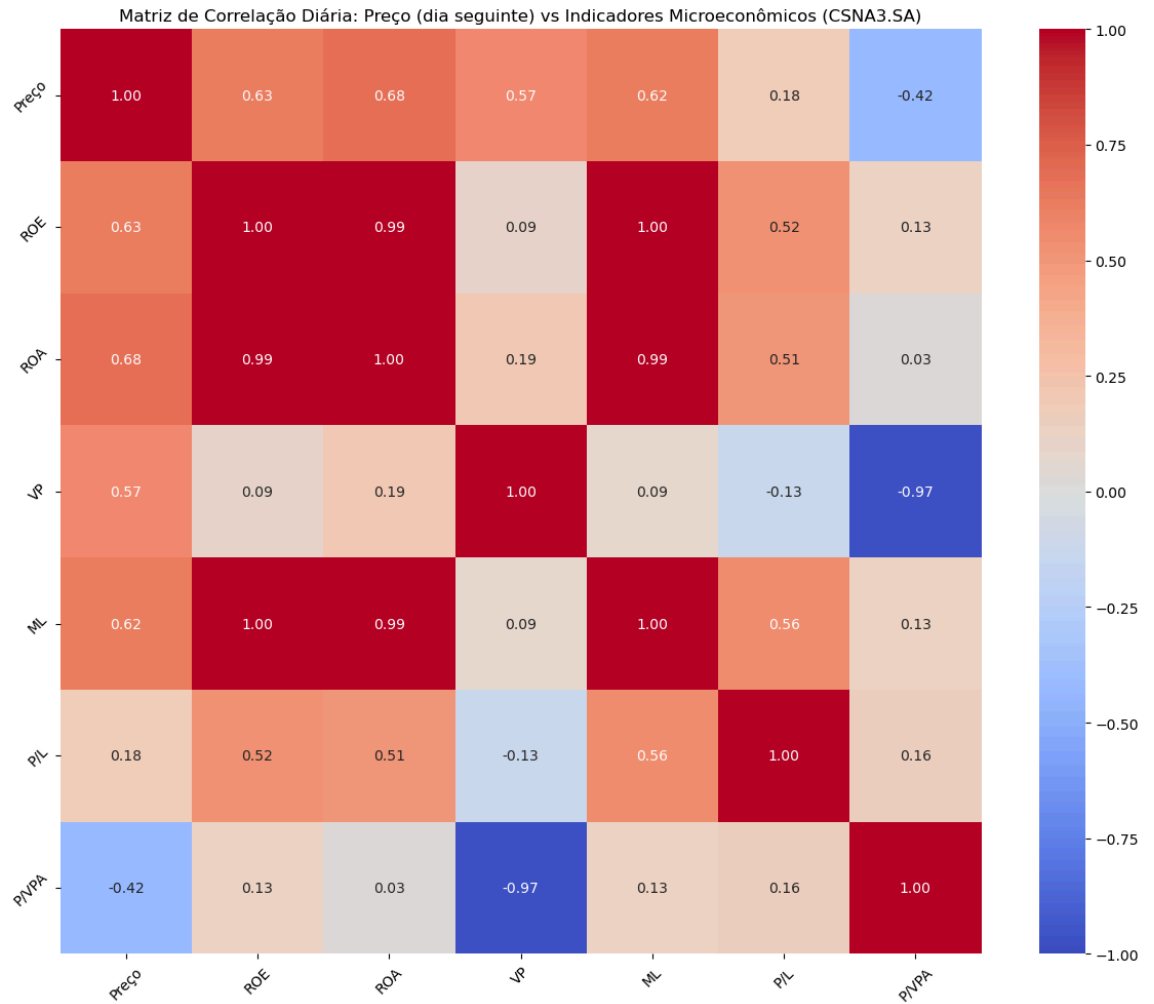
BEEF3.SA



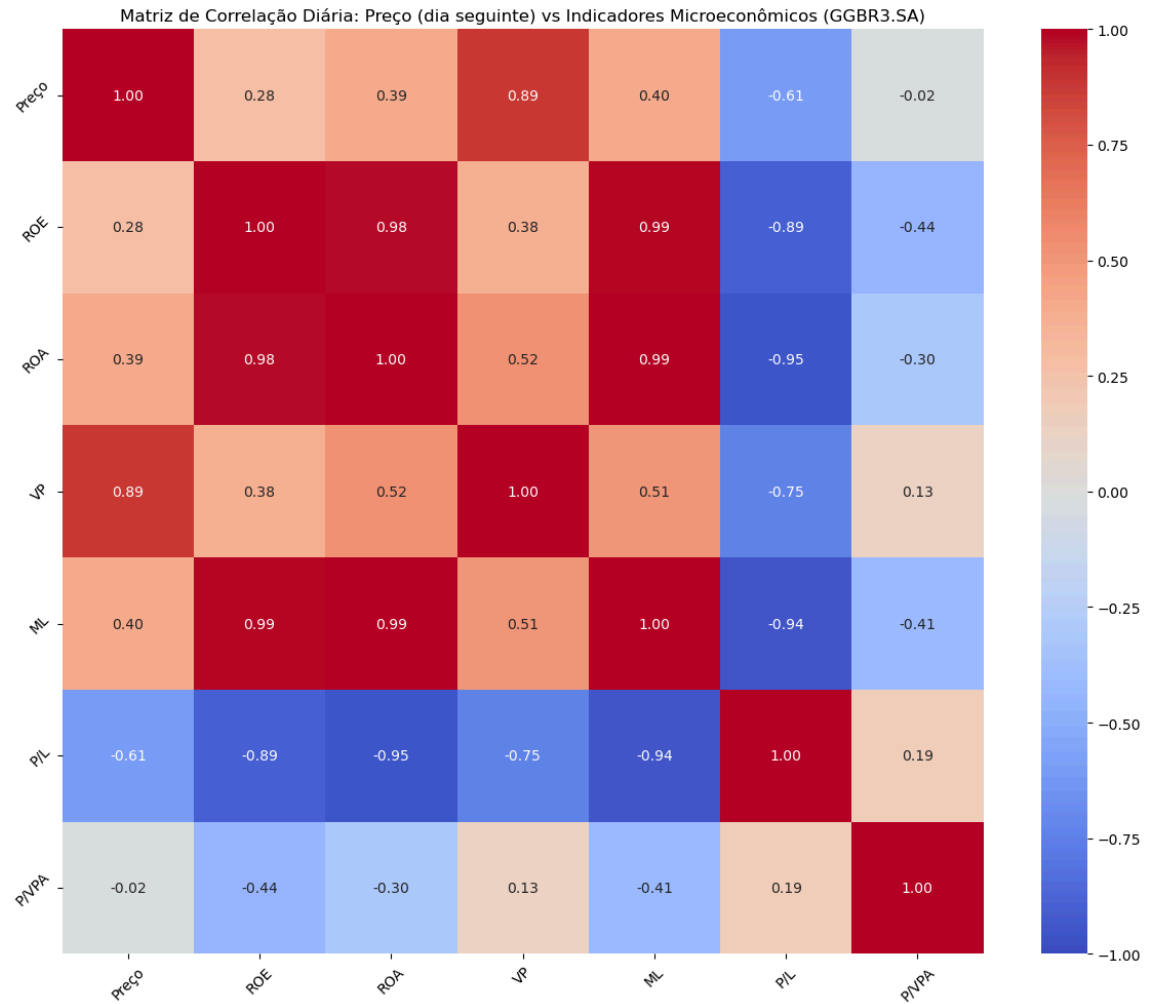
BRFS3.SA



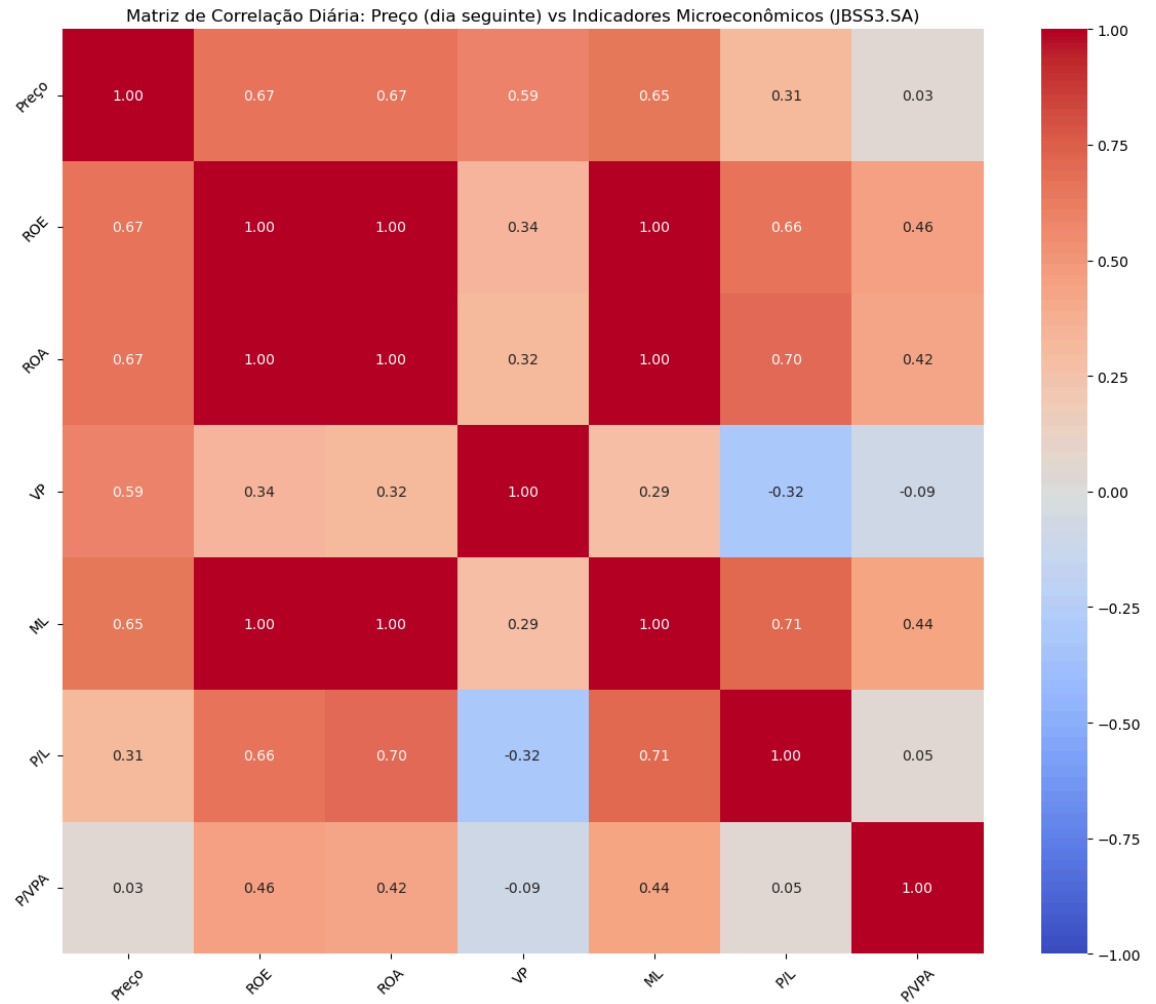
CSNA3.SA



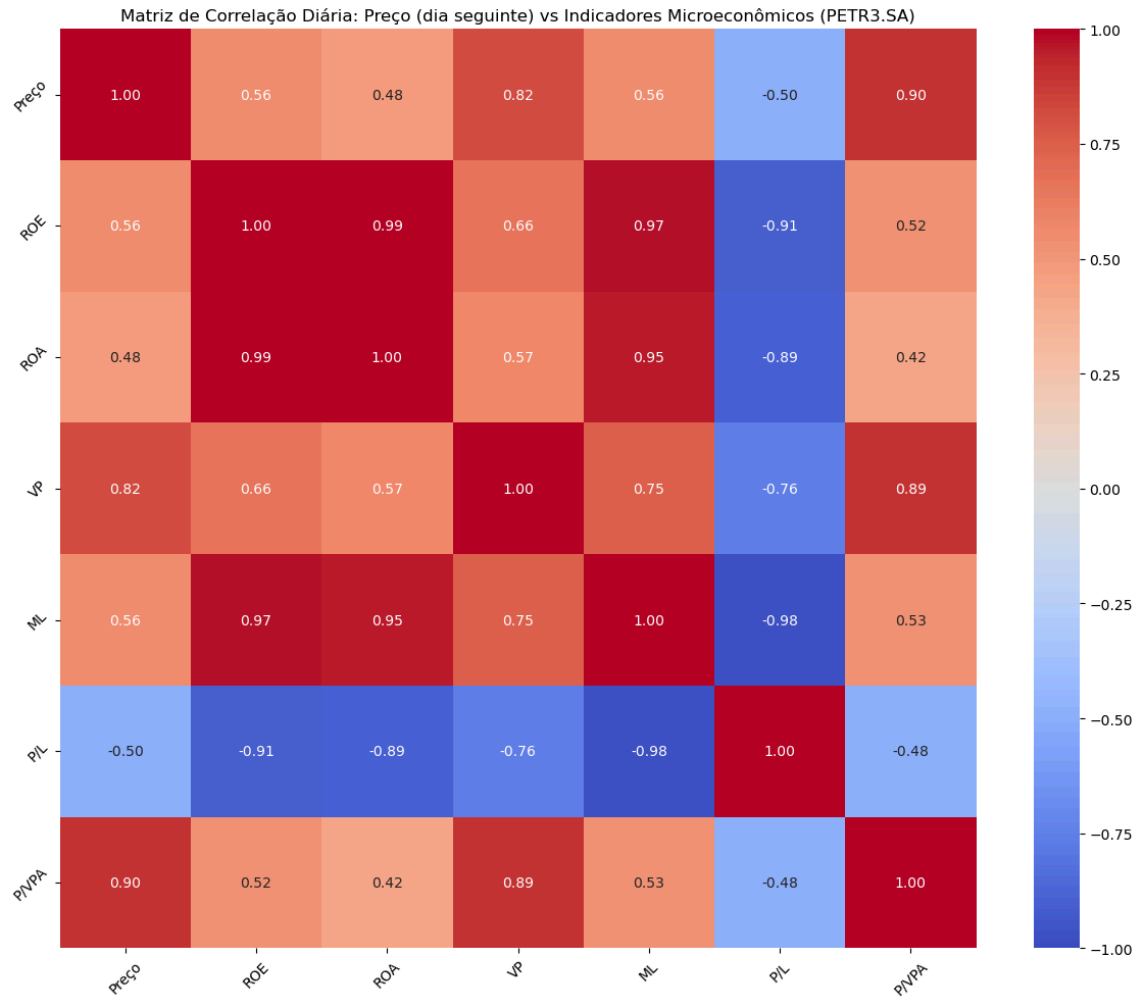
GGBR3.SA



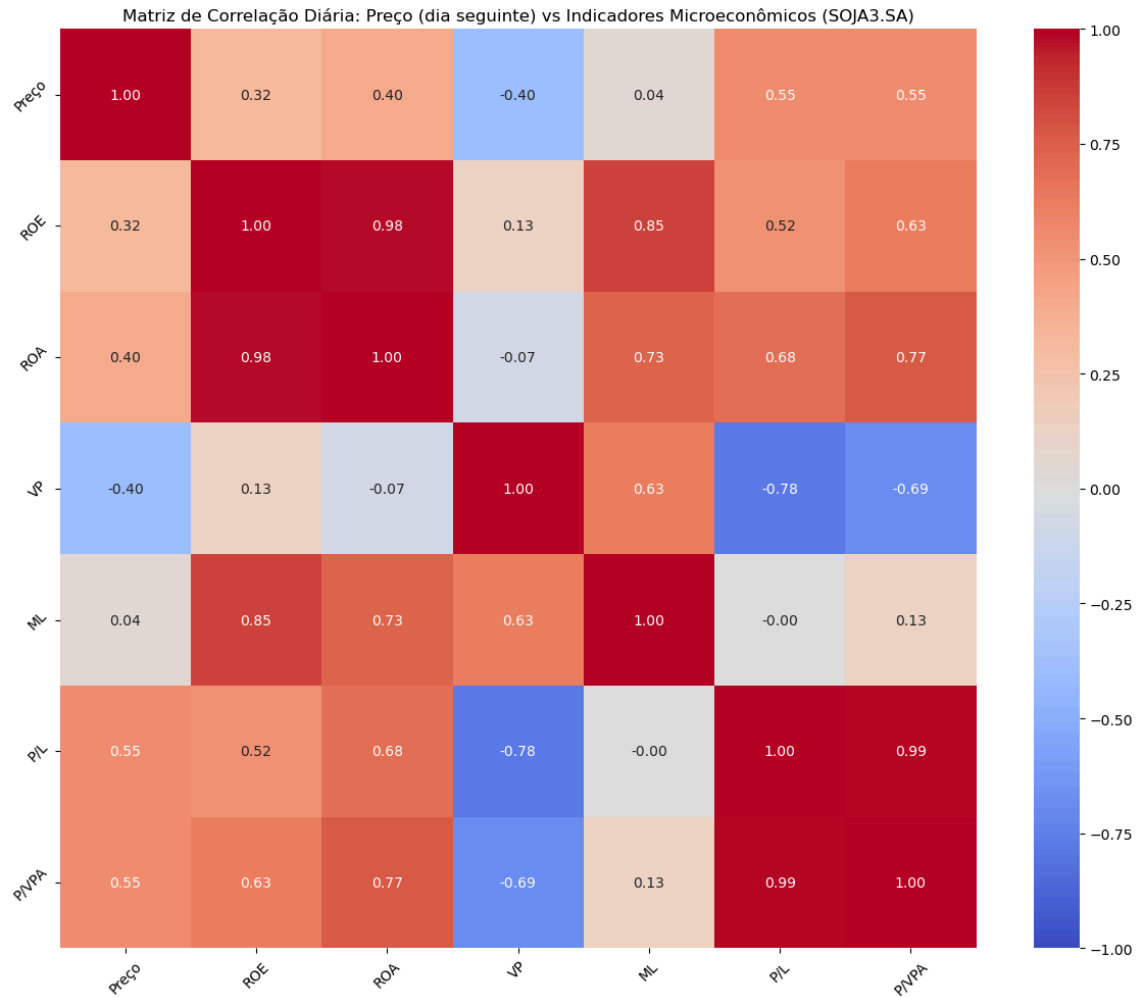
JBSS3.SA



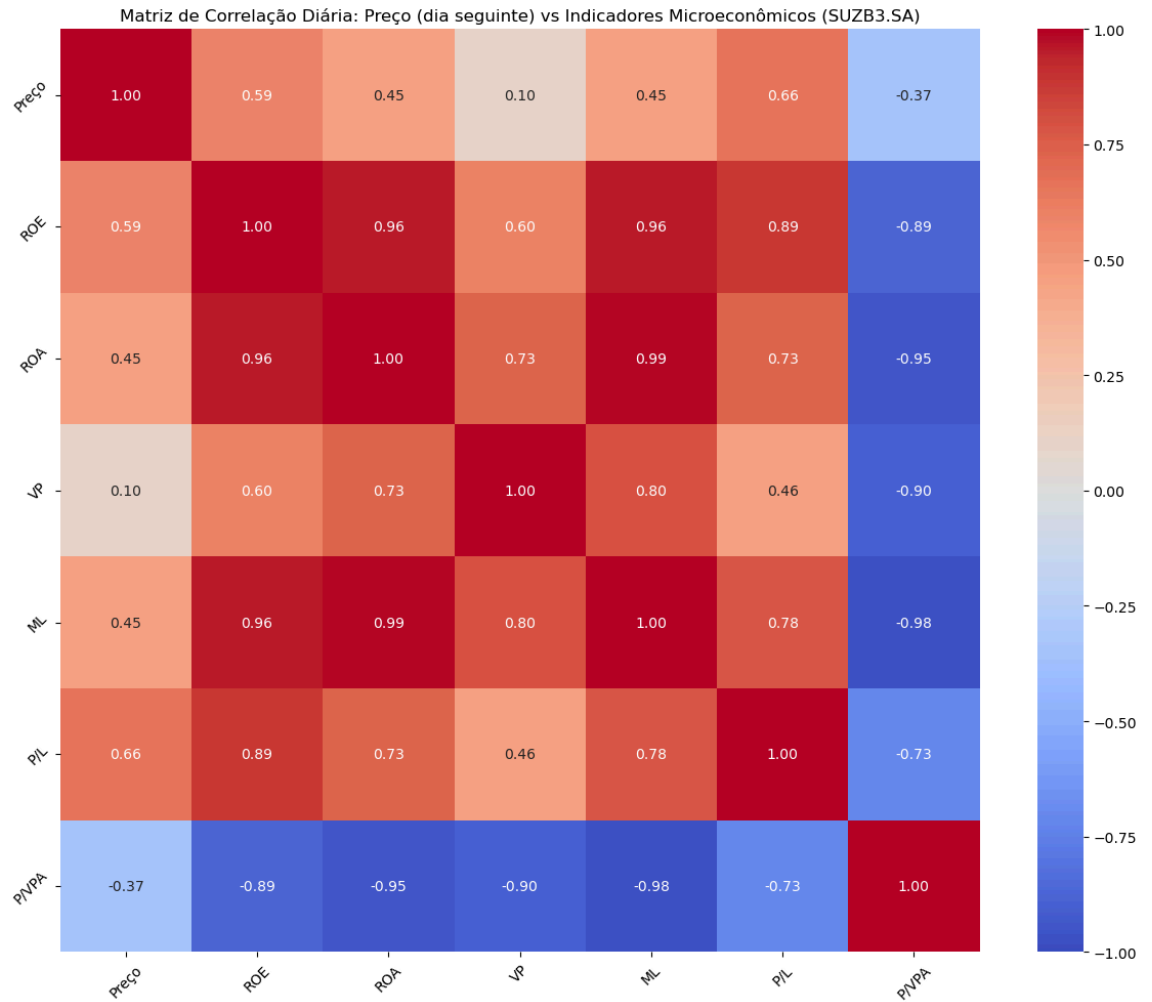
PETR3.SA



SOJA3.SA



SUZB3.SA



VALE3.SA

