

Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Ciências
Campos De Bauru

ANA CARLA CASERTA
LEANDRO GONZAGA

**MODELAGEM PREDITIVA E *FORECAST* SOBRE SÉRIES
TEMPORAIS DAS EXPORTAÇÕES BRASILEIRAS ATRAVÉS DA
NOMENCLATURA COMUM DO MERCOSUL (NCM) UTILIZANDO
*PROPHET***

BAURU
Junho/2023

ANA CARLA CASERTA
LEANDRO GONZAGA

**MODELAGEM PREDITIVA E *FORECAST* SOBRE SÉRIES
TEMPORAIS DAS EXPORTAÇÕES BRASILEIRAS ATRAVÉS DA
NOMENCLATURA COMUM DO MERCOSUL (NCM) UTILIZANDO
*PROPHET***

Trabalho de Conclusão de Curso do
Curso de Sistemas de Informação da
Universidade Estadual Paulista “Júlio
de Mesquita Filho”, Faculdade de
Ciências, Campus Bauru.
Orientador: Prof. João Pedro Albino

BAURU
Junho/2023

	Caserta, Ana Carla. Gonzaga Leandro
C338m	Modelagem preditiva e forecast sobre séries temporais das exportações brasileiras através da nomenclatura comum do mercosul (ncm) utilizando prophet / Ana Carla Caserta. Leandro Gonzaga. -- Bauru, 2023 40 p. : il., tabs. Trabalho de conclusão de curso (Bacharelado - Sistemas de Informação) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências, Bauru Orientador: João Pedro Albino 1. Análise de dados. 2. R. 3. Forecast. 4. R. 5. NCM. I. Título.

ANA CARLA CASERTA
LEANDRO GONZAGA

**MODELAGEM PREDITIVA E *FORECAST* SOBRE SÉRIES
TEMPORAIS DAS EXPORTAÇÕES BRASILEIRAS ATRAVÉS DA
NOMENCLATURA COMUM DO MERCOSUL (NCM) UTILIZANDO
*PROPHET***

Trabalho de Conclusão de Curso do
Curso de Sistemas de Informação da
Universidade Estadual Paulista "Júlio de
Mesquita Filho", Faculdade de Ciências,
Campus Bauru.

Bauru, ____ de _____ de ____.

BANCA EXAMINADORA

Prof. João Pedro Albino

Orientador

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Prof. Júlio César Melo

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Engenharia

Engenharia de Produção

RESUMO

Os dados estão mudando nosso mundo e a maneira como vivemos e trabalhamos em um ritmo sem precedentes. Mais do que nunca, habilidades e ferramentas de análise de negócios são essenciais para tomadas de decisões estratégicas, previsões de investimento e planejamento do negócio. Mensalmente o governo brasileiro fornece informações sobre todos os produtos exportados de acordo com sua NCM (Nomenclatura Comum do Mercosul). Essas informações são atualizadas em uma base on-line no domínio da Balança Comercial Brasileira. O presente trabalho visa demonstrar a tendência do volume de exportações do país ao longo do tempo e fazer previsões futuras de acordo com a sua NCM. Ao analisar os dados de exportação utilizando o NCM, é possível obter uma visão detalhada sobre os produtos comercializados, identificar tendências, padrões sazonais e variações ao longo do tempo. Essa análise pode ser útil para diversos propósitos, como tomada de decisões estratégicas, identificação de mercados potenciais, análise de concorrência, avaliação de desempenho e monitoramento do comércio internacional.

Palavras-Chave: Análise de Dados; Exportação; NCM, *RStudio*, *R*, *Prophet*, *Forecast* e *Flexdashboard* .

ABSTRACT

Data is changing our world and the way we live and work at an unprecedented pace. More than ever, business analysis skills and tools are essential for strategic decision-making, investment forecasting, and business planning. Monthly, the Brazilian government provides information about all exported products according to their NCM (Common Nomenclature of Mercosul). This information is updated online on the Brazilian Trade Balance domain. The present work aims to demonstrate the country's export volume trend over time and make future predictions based on its NCM. By analyzing export data using the NCM, it is possible to obtain a detailed view of the traded products, identify trends, seasonal patterns, and variations over time. This analysis can be useful for various purposes, such as strategic decision-making, identifying potential markets, analyzing competition, performance evaluation, and monitoring international trade.

Palavras-Chave: Data Analysis; Exportation; NCM, *RStudio*, *R*, *Prophet*, *Forecast* e *Flexdashboard* .

LISTA DE FIGURAS

Figura 1: Exemplo tabela NCM.....	9
Figura 2: Diagrama de Bibliotecas do R	17
Figura 3: Forecast para 2023.....	29
Figura 4: MAPE	30
Figura 5: Dashboard	32
Figura 6: Visualização do comportamento dos dados preditos para 2023.	33
Figura 7: Tendência anual dos valores apresentados.	33

LISTA DE SIGLAS E ABREVIACES

IDE	Ambiente de Desenvolvimento Integrado
NCM	Nomenclatura Comum do Mercosul
SH	Sistema Harmonizado de Designao e Codificao de Mercadorias
MSE	Erro Mdio Quadrtico
MAE	Erro Mdio Absoluto
MAPE	Valor Real

SUMÁRIO

1	INTRODUÇÃO	5
2	FUNDAMENTOS	7
2.1	Nomenclatura Comum Mercosul (NCM).....	7
2.2	Balança Comercial	9
2.3	<i>Forecast</i>	10
2.4	Importância do <i>Forecast</i> para o mercado financeiro	11
3	METODOLOGIA	14
3.1	Linguagem de Programação	14
3.1.1	Instalação do R	15
3.1.2	Biblioteca Prophet	17
3.1.3	Biblioteca Flexdashboard	20
3.1.4	Biblioteca Tidyverse	22
3.2	Ambiente de Desenvolvimento Integrado (IDE)	23
3.3	Extração e entendimento dos dados	25
3.4	Preparação e Limpeza	26
3.5	Modelagem.....	27
3.6	Análise.....	29
3.7	Apresentação	31
4	CONCLUSÃO	34
	REFERÊNCIAS	36

1 INTRODUÇÃO

A análise de dados de exportação por NCM (Nomenclatura Comum do Mercosul) no Brasil desempenha um papel fundamental na compreensão e monitoramento das atividades comerciais do país. A NCM é um sistema internacionalmente reconhecido de classificação de mercadorias que permite a padronização e identificação das diferentes categorias de produtos que são exportados e importados.

O Brasil é um dos maiores exportadores do mundo e possui uma economia diversificada, abrangendo setores como agricultura, indústria, mineração e serviços. Analisar os dados de exportação por NCM é essencial para compreender as principais áreas de atuação econômica do país, identificar tendências, avaliar a competitividade de setores específicos e tomar decisões estratégicas no comércio internacional.

Através da análise de dados de exportação por NCM, é possível identificar os produtos mais exportados pelo Brasil, entender os destinos dessas exportações e analisar a evolução ao longo do tempo. Essa análise pode revelar informações valiosas sobre a participação do Brasil em diferentes mercados globais, o desempenho de setores específicos da economia e as mudanças nas preferências dos compradores internacionais.

Essa análise também permite identificar oportunidades de diversificação de mercados e produtos, auxiliando empresas e formuladores de políticas a tomar decisões baseadas em dados. Por exemplo, a identificação de produtos com alto potencial de exportação pode direcionar investimentos e esforços para aumentar a produção e a competitividade nessas áreas.

Essa análise também é relevante para entender os efeitos de acordos comerciais, políticas governamentais e flutuações econômicas globais nas exportações brasileiras. Ao monitorar as tendências e os padrões de comércio, é possível antecipar desafios e identificar oportunidades para impulsionar o crescimento econômico e melhorar a posição do Brasil no comércio internacional.

Basicamente, a análise de dados de exportação por NCM no Brasil é uma ferramenta essencial para compreender a dinâmica do comércio exterior do

país. Ela fornece *insights* valiosos sobre os produtos mais exportados, os destinos de exportação, as tendências de mercado e as oportunidades de crescimento. Essa análise desempenha um papel fundamental no planejamento estratégico, na tomada de decisões e no monitoramento do desempenho econômico do Brasil no contexto global.

Ao longo desta monografia serão apresentados conceitos importantes para o entendimento da motivação e utilidade da análise desenvolvida neste trabalho. No capítulo 2 estão apresentados os fundamentos do que é a Nomenclatura Comum do Mercosul (NCM) e a sua aplicabilidade nas questões aduaneiras, bem como sua influência no entendimento da balança comercial do país. Ainda no capítulo 2, estão explanados os conceitos de *forecast*, e sua importância no mercado financeiro, e o que é a balança comercial. No capítulo 3 são apresentados os passos e tecnologias utilizadas para o desenvolvimento da análise proposta nesta monografia, explorando a linguagem de programação R e suas bibliotecas, tal qual o ambiente de desenvolvimento integrado, programa usado para compilação de todo o código. Por fim, a conclusão, elaborada após todo o processo de extração, limpeza, modelagem e análise dos dados.

2 FUNDAMENTOS

Neste capítulo será apresentado uma breve introdução sobre os assuntos tratados na dissertação, clarificando o que é e como funciona a classificação NCM, informações sobre previsões financeiras e a dinâmica de exportação de insumos no país, bem como, apresentação dos dados visto sob a ótica de *Business Intelligence* por meio das ferramentas de análise de dados e *Data Visualization*.

2.1 Nomenclatura Comum Mercosul (NCM)

A NCM é um sistema de classificação de mercadorias utilizado por diversos países, principalmente os membros do Mercosul, como Brasil, Argentina, Paraguai, Uruguai e Venezuela. Também conhecida como NCM/SH (Nomenclatura Comum do Mercosul e Sistema Harmonizado), essa classificação é baseada no Sistema Harmonizado de Designação e Codificação de Mercadorias (SH), desenvolvido pela Organização Mundial das Aduanas (OMA).

A NCM é composta por uma sequência de códigos numéricos que representam diferentes categorias de mercadorias. Esses códigos são padronizados internacionalmente e têm como objetivo facilitar o comércio internacional, harmonizando as informações sobre as mercadorias em diferentes países e permitindo a aplicação de impostos, tarifas e regulamentações comerciais de forma consistente.

Cada código NCM possui uma descrição que detalha a mercadoria correspondente. Essa descrição é utilizada para classificar as mercadorias de acordo com suas características, como natureza, função, composição e uso. A classificação correta das mercadorias na NCM é fundamental para a aplicação de tarifas aduaneiras, impostos de importação e exportação, bem como, para o cumprimento de regulamentações específicas de cada país.

A NCM é utilizada em diversos processos relacionados ao comércio internacional, tais como:

- Desembaraço aduaneiro: Ao importar ou exportar mercadorias, é necessário informar a NCM correta para que os órgãos aduaneiros possam aplicar as tarifas e impostos apropriados e verificar se as mercadorias estão em conformidade com as regulamentações.
- Estatísticas comerciais: A NCM é amplamente utilizada para a coleta e análise de dados sobre o comércio internacional. Ela permite o registro e a comparação de informações sobre importações e exportações de diferentes países, auxiliando na elaboração de políticas comerciais e no monitoramento do desempenho econômico.
- Regulamentações específicas: Alguns produtos estão sujeitos a regulamentações específicas de cada país, como normas de segurança, saúde, meio ambiente e controle de qualidade. A NCM é utilizada como base para identificar e classificar esses produtos, garantindo que eles atendam aos requisitos legais antes de serem comercializados.

É importante ressaltar que a classificação correta na NCM é uma tarefa complexa que requer conhecimento especializado e compreensão das características das mercadorias. Portanto, é comum que as empresas recorram a consultorias especializadas ou órgãos competentes para obter orientação na classificação adequada de suas mercadorias.

Para aprofundar o conhecimento sobre a NCM e suas aplicações, é possível consultar a legislação e as diretrizes aduaneiras específicas de cada país. No Brasil, por exemplo, a Secretaria de Comércio Exterior (SECEX), vinculada ao Ministério da Economia, é responsável pela administração da NCM e fornece informações detalhadas sobre sua utilização.

Segundo informações do site oficial da Receita Federal, a tabela de NCM funciona da seguinte forma: as mercadorias estão ordenadas sistematicamente na NCM, a priori, de forma progressiva, de acordo com o seu grau de elaboração, principiando pelos animais vivos e terminando com as obras de arte, passando por matérias-primas e produtos semi-acabados.

Assim, de modo geral, à medida que cresce a participação do homem na elaboração da mercadoria, mais elevado é o número do Capítulo em que ela

será classificada. Os seis primeiros dígitos da NCM seguem, por convenção internacional, o SH e seus dois últimos dígitos são definidos pelo Mercosul.

A NCM tem a seguinte estrutura:

- 6 Regras Gerais para Interpretação do Sistema Harmonizado e 2 Regras Gerais Complementares;
- Notas de Seção, de Capítulo, de Subposição e Complementares;
- Lista ordenada de códigos em níveis de posição (4 dígitos), subposição (5 e 6 dígitos), item (7 dígitos) e subitem (8 dígitos), distribuídos em 21 Seções e 96 Capítulos.

Figura 1: Exemplo tabela NCM

Tabela NCM (vigente em 29/06/2023) Última Atualização: 01/04/2023 - Ato Legal: Resolução Camex nº 440/2022		
Código	Descrição	Descrição Concatenada
01	Animais vivos.	01: Animais vivos.
01.01	Cavalos, asininos e muares, vivos.	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos.
0101.2	- Cavalos:	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos. 0101.2: - Cavalos:
0101.21.00	-- Reprodutores de raça pura	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos. 0101.2: - Cavalos: 0101.21.00: -- Reprodutores de raça pura
0101.29.00	-- Outros	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos. 0101.2: - Cavalos: 0101.29.00: -- Outros
0101.30.00	- Asininos	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos. 0101.30.00: - Asininos
0101.90.00	- Outros	01: Animais vivos. 01.01: Cavalos, asininos e muares, vivos. 0101.90.00: - Outros
01.02	Animais vivos da espécie bovina.	01: Animais vivos. 01.02: Animais vivos da espécie bovina.
0102.2	- Bovinos domésticos:	01: Animais vivos. 01.02: Animais vivos da espécie bovina. 0102.2: - Bovinos domésticos:
0102.21	-- Reprodutores de raça pura	01: Animais vivos. 01.02: Animais vivos da espécie bovina. 0102.2: - Bovinos domésticos: 0102.21: -- Reprodutores de raça pura
0102.21.10	Prenhes ou com cria ao pé	01: Animais vivos.

Fonte: <https://portalunico.siscomex.gov.br/classif/#/nomenclatura/tabela?perfil=publico>

2.2 Balança Comercial

A balança comercial é um conceito econômico que se refere à diferença entre o valor das exportações e o valor das importações de bens e serviços de

um país durante um determinado período de tempo, geralmente um ano. É um dos principais componentes da conta corrente de uma nação.

Quando o valor das exportações de um país é maior do que o valor das importações, diz-se que o país possui uma balança comercial positiva, superávit comercial ou excedente comercial. Isso significa que o país está exportando mais do que está importando, gerando uma entrada líquida de moeda estrangeira e uma fonte de receita para a economia.

Por outro lado, quando o valor das importações é maior do que o valor das exportações, ocorre um déficit comercial ou uma balança comercial negativa. Nesse caso, o país está importando mais do que está exportando, o que implica em uma saída líquida de moeda estrangeira e uma perda de receita para a economia.

A balança comercial é importante porque reflete a competitividade de um país no comércio internacional e seu grau de abertura econômica. Um superávit comercial pode indicar que um país está produzindo bens e serviços competitivos que são valorizados no mercado internacional. Além disso, um superávit comercial pode contribuir para o crescimento econômico, pois aumenta a demanda doméstica por bens e serviços produzidos localmente.

No entanto, um déficit comercial não é necessariamente uma indicação negativa, uma vez que pode refletir o desejo dos consumidores de adquirir bens e serviços importados. Também pode ser influenciado por fatores como a estrutura econômica do país, a taxa de câmbio e as políticas comerciais adotadas.

A balança comercial é monitorada pelos governos, economistas e analistas de mercado para avaliar o desempenho econômico de um país, sua competitividade e sua dependência em relação ao comércio internacional. É um indicador importante para o planejamento de políticas econômicas, decisões de investimento e avaliação da saúde geral da economia.

2.3 *Forecast*

Forecast, ou previsão, é o processo de estimar ou prever eventos futuros com base em informações e dados disponíveis no presente. É uma técnica

amplamente utilizada em diversas áreas, como economia, negócios, meteorologia, finanças e análise de séries temporais.

O objetivo do é fornecer uma visão do que é esperado no futuro, permitindo que indivíduos, empresas e organizações tomem decisões informadas e planejem suas atividades com base nessas previsões. O pode ser aplicado para prever uma ampla variedade de elementos, tais como demanda de mercado, vendas, receitas, despesas, condições climáticas, preços de commodities, entre outros.

Existem diferentes abordagens e métodos para realizar previsões, incluindo técnicas estatísticas, modelos econométricos, modelos de séries temporais, aprendizado de máquina e análise de tendências históricas. Cada método possui suas próprias suposições e requisitos de dados, e a escolha do método depende do tipo de dado disponível, do horizonte de previsão desejado e da natureza do fenômeno que está sendo previsto.

É importante ressaltar que as previsões não são predições exatas do futuro, mas sim estimativas baseadas em informações e análises disponíveis. As incertezas e imprevisibilidades inerentes aos eventos futuros podem levar a desvios em relação às previsões realizadas. No entanto, mesmo com essa limitação, as previsões são valiosas para orientar a tomada de decisões, fornecer *insights* estratégicos e auxiliar no planejamento e gerenciamento de atividades.

As previsões podem ser feitas para diferentes horizontes de tempo, desde curto prazo (dias, semanas) até longo prazo (anos, décadas). É comum que as previsões sejam atualizadas periodicamente à medida que novos dados se tornam disponíveis e a situação evolua.

Em resumo, o é o processo de estimar ou prever eventos futuros com base em informações e dados disponíveis no presente. Ele desempenha um papel fundamental em várias áreas, fornecendo informações valiosas para a tomada de decisões, planejamento estratégico e gerenciamento de riscos.

2.4 Importância do *Forecast* para o mercado financeiro

Fazer previsões financeiras é de extrema importância para diversas razões. Aqui estão algumas das principais razões pelas quais o *forecast* financeiro é importante:

- Planejamento estratégico: O financeiro permite que as empresas desenvolvam um plano estratégico de curto, médio e longo prazo. Com base nas previsões financeiras, as empresas podem estabelecer metas e objetivos realistas, identificar oportunidades de crescimento e tomar decisões informadas para alcançar resultados desejados.
- Tomada de decisões: O *forecast* financeiro fornece informações valiosas para a tomada de decisões empresariais. Ele ajuda a identificar tendências e padrões, antecipar possíveis desafios e avaliar o impacto financeiro de diferentes cenários. Com base nas previsões, os gestores podem tomar decisões estratégicas relacionadas a investimentos, alocação de recursos, expansão de negócios e outras atividades que afetam a saúde financeira da organização.
- Gerenciamento de fluxo de caixa: O *forecast* financeiro é essencial para o gerenciamento eficiente do fluxo de caixa. Ele permite que as empresas prevejam suas receitas e despesas futuras, identifiquem períodos de escassez de caixa e tomem medidas proativas para garantir a disponibilidade de fundos quando necessário. Isso é especialmente crítico para pequenas e médias empresas, que podem enfrentar desafios de liquidez e dependem de uma gestão rigorosa do fluxo de caixa para sua sobrevivência.
- Planejamento de investimentos: O *forecast* financeiro desempenha um papel fundamental no planejamento de investimentos. Ele ajuda a determinar a viabilidade e a rentabilidade de projetos de investimento, permitindo que as empresas identifiquem oportunidades lucrativas e avaliem os riscos associados. As previsões financeiras também auxiliam na definição de estratégias de financiamento para suportar esses investimentos.

- Monitoramento de desempenho: As previsões financeiras servem como uma linha de base para o monitoramento e controle do desempenho financeiro real em relação ao planejado. Ao comparar os resultados reais com as previsões, as empresas podem identificar desvios, analisar as causas subjacentes e tomar medidas corretivas para manter o desempenho financeiro em linha com as expectativas.

Em síntese, o financeiro é crucial para o planejamento estratégico, tomada de decisões informadas, gerenciamento do fluxo de caixa, planejamento de investimentos e monitoramento do desempenho financeiro. Ao fornecer uma visão do futuro financeiro da empresa, as previsões financeiras capacitam as organizações a tomar medidas proativas e estratégicas para alcançar o sucesso sustentável.

3 METODOLOGIA

Para o desenvolvimento da análise e apresentação dos dados mostrados nesta monografia, serão utilizadas ferramentas em quatro grandes campos do trabalho com grande volume de dados: extração, análise, tratamento e apresentação dos dados.

3.1 Linguagem de Programação

A linguagem R é uma poderosa linguagem de programação e ambiente de software utilizado para análise de dados, estatística e visualização. Criada por Ross Ihaka e Robert Gentleman na década de 1990, ela se tornou uma ferramenta amplamente adotada pela comunidade estatística e científica devido à sua flexibilidade, capacidade de manipulação de dados e extensibilidade (TIPPMANN, 2015).

Uma das principais características do R é a sua vasta coleção de pacotes e bibliotecas. Esses pacotes oferecem uma ampla gama de funcionalidades adicionais que estendem a linguagem base do R, permitindo a realização de análises estatísticas complexas, visualizações de dados avançadas e a implementação de algoritmos especializados em diversas áreas.

A linguagem R suporta uma variedade de técnicas estatísticas, como regressão linear e logística, análise de séries temporais, modelos de mistura, análise de sobrevivência, entre outros. Além disso, ela oferece recursos para realizar testes estatísticos, construir modelos preditivos e realizar simulações.

A visualização dos resultados é uma parte essencial da análise de dados, e o R possui recursos gráficos poderosos. O pacote ggplot2, por exemplo, permite criar gráficos estatísticos elegantes e personalizados com facilidade. Com a capacidade de criar gráficos estáticos e interativos, o R oferece uma maneira eficaz de explorar, comunicar e visualizar informações a partir dos dados.

Outra vantagem do R é a sua capacidade de trabalhar com grandes conjuntos de dados. Através de técnicas de programação vetorial, é possível manipular e processar grandes volumes de dados de forma eficiente. Além disso,

o R possui recursos para lidar com diferentes formatos de dados, como planilhas, bancos de dados, arquivos CSV e até mesmo a integração com a web.

A comunidade em torno do R é ativa e colaborativa, o que resulta em uma vasta quantidade de recursos disponíveis. Existem fóruns, listas de discussão, grupos de usuários e repositórios online onde os usuários podem compartilhar código, solucionar problemas e colaborar em projetos. Essa comunidade também contribui para o desenvolvimento e aprimoramento contínuo do R, com a criação de novos pacotes e melhorias na linguagem.

A reprodutibilidade é uma parte fundamental do processo de análise de dados, e o R oferece recursos para garantir a replicabilidade dos resultados. Ao usar scripts e notebooks, é possível documentar e compartilhar todo o processo analítico, desde a importação dos dados até a criação de gráficos e modelos, garantindo que os resultados possam ser reproduzidos no futuro.

3.1.1 Instalação do R

A instalação do R é um passo essencial para começar a usar essa linguagem de programação e ambiente estatístico. Aqui estão algumas informações sobre o processo de instalação do R:

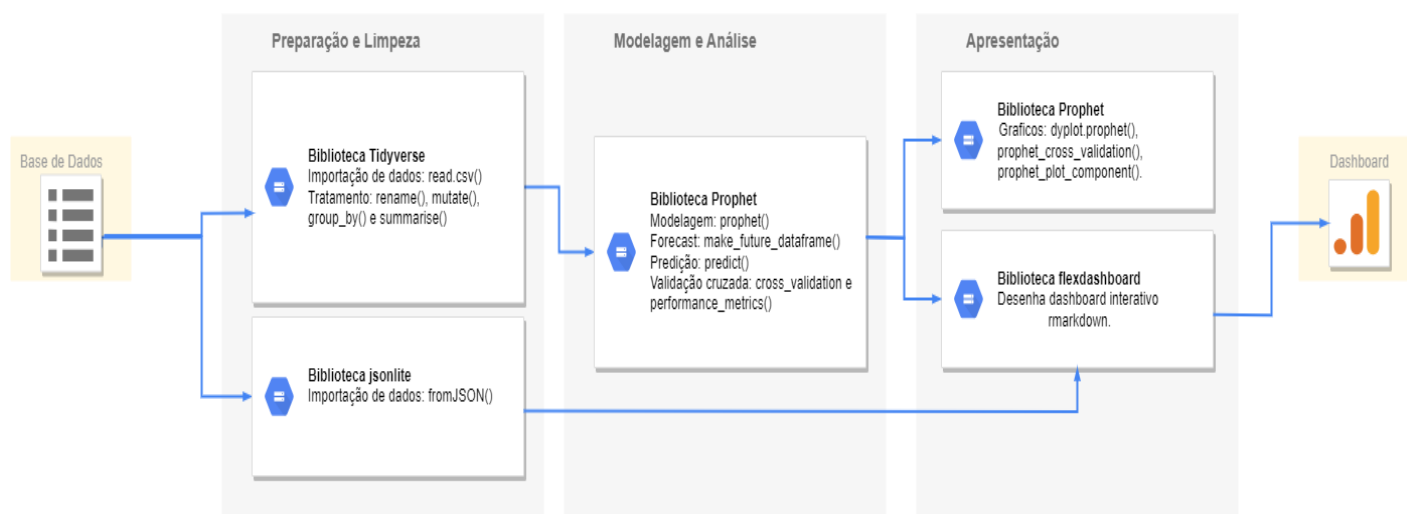
- **Baixando o instalador:** Para começar, você precisa acessar o site oficial do R Project (<https://www.r-project.org/>) e localizar a seção de downloads. Lá você encontrará versões específicas para diferentes sistemas operacionais, como Windows, macOS e Linux. Selecione a versão compatível com o seu sistema operacional.
- **Escolhendo um espelho:** Ao clicar no link de download correspondente ao seu sistema operacional, você será redirecionado para uma página que lista vários espelhos (*mirrors*). Os espelhos são sites que hospedam o arquivo de instalação do R. Escolha um espelho próximo à sua localização geográfica para garantir uma velocidade de download mais rápida.
- **Baixando e executando o instalador:** No espelho selecionado, procure pelo link de download do instalador do R e faça o download do arquivo. O

nome do arquivo pode variar dependendo do sistema operacional. Após o download, execute o instalador para iniciar o processo de instalação.

- Configurações durante a instalação: Durante o processo de instalação, você será solicitado a fazer algumas configurações. Essas configurações podem variar dependendo do sistema operacional, mas geralmente incluem:
 - Seleção do idioma: Escolha o idioma desejado para a instalação.
 - Localização da instalação: Escolha o diretório onde o R será instalado em seu computador. O local padrão geralmente é recomendado, a menos que você tenha uma preferência específica.
 - Componentes a serem instalados: O instalador do R pode oferecer a opção de escolher componentes adicionais para instalar, como documentação, pacotes sugeridos e exemplos. Você pode selecionar as opções de acordo com suas necessidades.
- Conclusão da instalação: Após fazer as configurações necessárias, o instalador do R prosseguirá com a instalação dos arquivos. Aguarde até que o processo seja concluído.
- Verificando a instalação: Após a instalação, você pode verificar se o R foi instalado corretamente. Abra o R executando o programa correspondente ao seu sistema operacional (por exemplo, "R" ou "RGui"). Isso abrirá a interface de linha de comando do R, onde você pode digitar comandos e executar análises estatísticas.

É importante ressaltar que esses são apenas os passos básicos para instalar o R. Suas funcionalidades são implementadas com a instalação e importação de bibliotecas. Na Figura 1 o diagrama define as principais bibliotecas utilizadas em cada etapa desse trabalho, o relacionamento entre elas e suas funções:

Figura 2: Diagrama de Bibliotecas do R



Fonte: Autor

3.1.2 Biblioteca Prophet

A biblioteca *Prophet* no R é uma ferramenta poderosa para realizar previsões de séries temporais. Desenvolvida pelo Facebook, ela foi projetada para facilitar o processo de modelagem e previsão de dados temporais, especialmente para usuários sem experiência em estatística avançada. A biblioteca *Prophet* oferece uma abordagem automatizada e flexível, permitindo que os usuários aproveitem os benefícios das técnicas de previsão de séries temporais de maneira eficiente e eficaz.

Uma das principais vantagens de usar a biblioteca *Prophet* é sua capacidade de lidar com uma ampla variedade de padrões e tendências comuns em séries temporais. Ela é capaz de lidar com sazonalidades diárias, semanais, mensais e anuais, bem como com feriados e outros eventos importantes. Isso permite que os usuários capturem e modelagem adequadamente os padrões sazonais em seus dados, resultando em previsões mais precisas e confiáveis.

Outra vantagem significativa é a facilidade de uso da biblioteca *Prophet*. Ela fornece uma interface intuitiva que simplifica todo o processo de previsão. Com apenas algumas linhas de código, é possível importar os dados, ajustar o modelo e gerar previsões futuras. Isso torna a biblioteca adequada tanto para

usuários iniciantes quanto para especialistas em séries temporais, agilizando o desenvolvimento de modelos e permitindo que os resultados sejam obtidos rapidamente.

A biblioteca *Prophet* também se destaca em sua capacidade de lidar com valores ausentes e *outliers* nos dados de séries temporais. Ela oferece métodos robustos de tratamento de dados faltantes e técnicas eficientes para lidar com observações anômalas. Isso é particularmente útil em cenários do mundo real, onde os dados podem ser imperfeitos e sujeitos a erros ou lacunas.

Além disso, a biblioteca *Prophet* possui recursos avançados de diagnóstico e avaliação de modelos. Ela permite que os usuários visualizem e interpretem facilmente os resultados das previsões, oferecendo gráficos e métricas relevantes para avaliar o desempenho do modelo. Isso facilita a compreensão das previsões e a tomada de decisões informadas com base nos resultados obtidos.

Em resumo, a biblioteca *Prophet* no R é uma ferramenta valiosa para a previsão de séries temporais. Suas vantagens incluem a capacidade de lidar com padrões complexos, facilidade de uso, tratamento de dados ausentes e *outliers*, e recursos avançados de diagnóstico e avaliação de modelos. Com essas vantagens, a biblioteca *Prophet* se torna uma opção atraente para aqueles que desejam realizar previsões precisas e escaláveis em séries temporais.

3.1.2.1 Funcionamento *Prophet*

O funcionamento da biblioteca *Prophet* baseia-se em um modelo de decomposição aditiva, que descreve a série temporal como uma combinação de tendência, sazonalidade e efeitos de feriados. Essa abordagem permite capturar diferentes padrões presentes nos dados históricos e gerar previsões com maior acuracidade para o futuro.

A biblioteca *Prophet* automatiza grande parte do processo de modelagem, tornando-o acessível para usuários com diferentes níveis de experiência. Para utilizar a biblioteca, geralmente são necessárias as seguintes etapas:

- Preparação dos dados: Os dados de série temporal devem ser organizados em um dataframe com duas colunas: "ds" (representando a data) e "y" (representando a variável alvo a ser prevista).
- Criação do modelo: Um objeto *Prophet* é criado e ajustado aos dados históricos. Durante esse processo, a biblioteca analisa os padrões temporais presentes na série, como tendência e sazonalidade, e ajusta o modelo a eles.
- Especificação de componentes adicionais: Além dos componentes padrão (tendência, sazonalidade), é possível adicionar componentes adicionais ao modelo, como efeitos de feriados ou eventos específicos.
- Geração de previsões: Com o modelo ajustado, é possível gerar previsões para um período futuro específico. A biblioteca calcula automaticamente os intervalos de confiança das previsões, fornecendo informações sobre a incerteza associada.
- Avaliação e visualização: A biblioteca permite visualizar as previsões geradas, incluindo a tendência, sazonalidade e intervalos de confiança. Isso ajuda a avaliar a qualidade das previsões e entender melhor os padrões subjacentes nos dados.

O funcionamento da biblioteca *Prophet* é baseado em algoritmos avançados, como regressão em componentes aditivos (ARIMA) e modelos Bayesianos. No entanto, a complexidade desses algoritmos é encapsulada pela biblioteca, permitindo que os usuários aproveitem seus benefícios sem precisar se preocupar com os detalhes técnicos.

A função de previsão do *Prophet* é definida como:

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

- $g(t)$ representa a componente de tendência, que modela as mudanças não periódicas no valor da série temporal ao longo do tempo. É uma função suave que captura o comportamento geral da série temporal.

- $s(t)$ representa a componente de sazonalidade, que captura as variações periódicas da série temporal, como padrões semanais, mensais ou anuais. O *Prophet* utiliza uma abordagem aditiva para modelar a sazonalidade, considerando a soma dos efeitos das diferentes sazonalidades presentes nos dados.
- $h(t)$ representa a componente de feriados, que leva em consideração os efeitos dos feriados na série temporal. Esses feriados podem ocorrer em horários irregulares ao longo de um ou mais dias. O *Prophet* permite fornecer um DataFrame com as datas e nomes dos feriados relevantes para que o modelo ajuste automaticamente os efeitos desses feriados nos padrões dos dados.
- $\varepsilon(t)$ é o termo de erro ou resíduo, que representa quaisquer mudanças idiossincráticas não explicadas pelas outras componentes do modelo. O *Prophet* assume que o termo de erro segue uma distribuição normal.

Em resumo, a função de previsão do *Prophet* é composta por uma componente de tendência ($g(t)$) que modela as mudanças de longo prazo, uma componente de sazonalidade ($s(t)$) que captura os padrões periódicos e uma componente de feriados ($h(t)$) que considera os efeitos dos feriados. O termo de erro ($\varepsilon(t)$) representa as variações não explicadas pelo modelo. O *Prophet* estima os parâmetros dessas componentes durante o ajuste do modelo aos dados observados, permitindo previsões futuras com base nessas componentes.

3.1.3 Biblioteca Flexdashboard

A biblioteca *flexdashboard* é uma extensão do pacote RMarkdown que permite criar painéis de controle interativos e *dashboards* flexíveis. Ela fornece uma maneira eficiente e fácil de compartilhar análises de dados interativas e visualizações com outras pessoas.

O *flexdashboard* baseia-se na sintaxe Markdown, o que o torna bastante acessível para aqueles que já estão familiarizados com essa linguagem de marcação. Ele permite a combinação de elementos estáticos e interativos em um único documento, permitindo a criação de painéis de controle dinâmicos e personalizados.

A principal vantagem do *flexdashboard* é a sua capacidade de criar dashboards responsivos, que se adaptam automaticamente a diferentes tamanhos de tela. Isso significa que os dashboards podem ser facilmente visualizados em dispositivos móveis, tablets e computadores, oferecendo uma experiência consistente em qualquer dispositivo.

A biblioteca também oferece diversos layouts e opções de personalização, permitindo que você crie dashboards com uma aparência profissional e adaptada às suas necessidades. É possível criar painéis com várias seções, colunas e diferentes tipos de gráficos e visualizações interativas, como tabelas, gráficos de linhas, gráficos de barras, mapas e muito mais.

Além disso, o *flexdashboard* suporta a interatividade por meio do uso de *widgets*, que permitem ao usuário interagir com os dados e visualizações. É possível adicionar controles deslizantes, botões, caixas de seleção e outros elementos interativos para filtrar, agrupar ou explorar os dados de maneira dinâmica.

Uma característica interessante do *flexdashboard* é a possibilidade de incorporar código R e criar visualizações dinâmicas diretamente no documento. Isso permite que você realize análises complexas, execute modelos estatísticos e gere gráficos atualizados automaticamente quando os dados são alterados.

A biblioteca *flexdashboard* é amplamente utilizada por analistas de dados, cientistas de dados e profissionais que desejam compartilhar análises interativas de forma rápida e eficiente. Sua versatilidade e facilidade de uso a tornaram uma escolha popular para a criação de *dashboards* para apresentações, relatórios, monitoramento de dados e muito mais.

3.1.4 Biblioteca Tidyverse

A biblioteca Tidyverse é um conjunto abrangente de pacotes desenvolvidos para a linguagem de programação R, que visam facilitar a manipulação, visualização e análise de dados. Ele foi projetado para fornecer uma abordagem coerente e eficiente para o trabalho com dados, oferecendo uma sintaxe consistente e intuitiva.

O Tidyverse consiste em vários pacotes, cada um com sua função específica, mas todos compartilhando princípios e filosofias comuns. Alguns dos pacotes mais conhecidos e amplamente utilizados incluem:

- *dplyr*: O pacote dplyr fornece um conjunto de ferramentas para manipulação de dados, como filtragem, seleção de colunas, ordenação, agrupamento e agregação. Ele utiliza uma sintaxe clara e concisa, permitindo que os usuários escrevam códigos mais legíveis e eficientes para manipular dados.
- *ggplot2*: Este pacote é amplamente reconhecido por sua capacidade de criar gráficos estatísticos elegantes e de alta qualidade. Ele segue a filosofia de "gramática dos gráficos", onde os gráficos são construídos a partir de camadas, permitindo uma personalização detalhada de cada elemento do gráfico.
- *tidyr*: O pacote tidyr oferece funções para transformar e reorganizar conjuntos de dados. Ele facilita tarefas como separar colunas em variáveis distintas, espalhar dados de uma coluna para várias colunas ou reunir várias colunas em uma única coluna.
- *purrr*: Este pacote fornece uma abordagem consistente para aplicar funções em listas e vetores. Ele oferece ferramentas para iteração, mapeamento de funções, filtragem e outras operações relacionadas à programação funcional.

- *readr*: O pacote *readr* é usado para importar dados de diferentes formatos, como arquivos CSV, Excel e texto delimitado. Ele possui funções eficientes para ler e interpretar dados, tornando o processo de importação mais rápido e confiável.

A principal vantagem do Tidyverse é a consistência da sintaxe entre os pacotes, o que facilita a aprendizagem e o uso eficiente das ferramentas. Além disso, o Tidyverse enfatiza a organização e a limpeza dos dados, seguindo os princípios do "tidy data", onde os dados são organizados em colunas, variáveis e observações de maneira estruturada e coerente.

A abordagem do Tidyverse permite que os usuários foquem na análise dos dados em vez de se preocuparem com as complexidades da manipulação de dados. Ele promove um fluxo de trabalho mais eficiente e produtivo, permitindo que os analistas de dados realizem tarefas complexas de forma rápida e precisa.

A documentação do Tidyverse é amplamente disponível, com recursos online, tutoriais e exemplos práticos que ajudam os usuários a se familiarizarem com os diferentes pacotes e suas funcionalidades. A comunidade R é bastante ativa e oferece suporte contínuo para aqueles que utilizam o Tidyverse em seus projetos de análise de dados.

Basicamente, o Tidyverse é uma biblioteca poderosa e versátil para análise de dados em R.

3.2 Ambiente de Desenvolvimento Integrado (IDE)

O IDE utilizado no presente trabalho é RStudio. O RStudio é um ambiente de desenvolvimento integrado (IDE, na sigla em inglês) para a linguagem de programação R. Ele fornece uma interface amigável e eficiente para escrever, executar e depurar código R. O RStudio é projetado especificamente para facilitar o trabalho com o R, oferecendo recursos avançados que aumentam a produtividade dos usuários.

Algumas das principais características e funcionalidades do RStudio incluem:

- Editor de código: O RStudio possui um editor de código R completo, com recursos como realce de sintaxe, auto completar, identificação automática e outras funcionalidades para facilitar a escrita do código.
- Console interativo: O RStudio oferece um console interativo, onde é possível executar comandos R em tempo real e visualizar os resultados imediatamente. Isso permite uma análise iterativa e exploratória dos dados.
- Gerenciamento de projetos: O RStudio permite a organização de projetos em um ambiente integrado. Ele oferece recursos para criar, abrir e gerenciar projetos, facilitando o controle de versões e a colaboração em equipe.
- Visualização de dados: O RStudio inclui recursos gráficos poderosos para a visualização de dados, permitindo criar gráficos interativos, personalizáveis e de alta qualidade diretamente no ambiente de trabalho.
- Gerenciamento de pacotes: O RStudio facilita a instalação, atualização e gerenciamento de pacotes R. Ele oferece uma interface intuitiva para encontrar, instalar e carregar pacotes necessários para análises específicas.
- Depuração de código: O RStudio possui recursos de depuração integrados que permitem identificar e corrigir erros no código R. Os usuários podem definir pontos de interrupção, examinar variáveis e percorrer o código passo a passo para entender o comportamento do programa.
- Integração com o Git: O RStudio possui integração nativa com o sistema de controle de versões Git. Isso permite o gerenciamento eficiente do

código-fonte, o controle de versões e a colaboração em projetos compartilhados.

Essas são apenas algumas das funcionalidades mais importantes do RStudio. Ele é amplamente utilizado pela comunidade de usuários do R devido à sua interface amigável, recursos avançados e suporte contínuo da equipe de desenvolvimento. O RStudio está disponível em versões gratuitas e pagas, atendendo tanto a usuários iniciantes quanto a profissionais avançados.

3.3 Extração e entendimento dos dados

Nessa etapa, foi realizada a coleta dos dados relevantes para o estudo, provenientes de fontes diversas. Além disso, foi dedicado um esforço para compreender profundamente esses dados, identificando sua origem, significado e estrutura. Isso nos permitirá ter uma visão abrangente do conjunto de dados e obter insights preliminares para direcionar as próximas etapas. A extração foi feita de forma manual, utilizando-se basicamente de duas bases de dados contendo os dados necessários para o modelo. A primeira base, composta pelo montante de informações do volume de exportações brasileiras entre os anos de 1997 à 2022, contém os dados pertinentes ao estudo e sazonalidade de entradas mensais. Abaixo têm-se as principais características:

- . Trata-se de uma base em formato “*tidy*”, ou seja, apresentam as variáveis dispostas em colunas e as observações em linhas. Esse formato também é conhecido como formato longo, e necessário para utilização de funções de tratamento de dados pela biblioteca *tidyverse*.

- . O arquivo está no formato csv, separado por ponto e vírgula. Logo sua importação foi feita pela função `read.csv()` da própria biblioteca *tidyverse*.

- . As principais variáveis que irão compor o processo de modelagem são: CO_ANO, na qual estão presentes as referências anuais de cada entrada do banco de dados; CO_MES, presentes as referências mensais de cada entrada do mesmo banco e VL_FOB, presentes as informações principais desse estudo que são os montantes em valores de moeda estrangeira (USD) por entrada registrada na base de dados.

Essa base foi utilizada diretamente na etapa de modelagem. Uma segunda base, contém as informações que trás referência ao código da NCM e a sua explicação por subdivisão. Essa foi utilizada na etapa de apresentação dos dados, no *dashboard* gerado para análise. Segue as principais características:

- . Devido ao fato dessa base não entrar no processo de modelagem, os dados não precisam ser tratados e sua estrutura JSON foi mantida.

- . Para sua importação foi utilizada a função *fromJSON()* da biblioteca *jsonlite*.

3.4 Preparação e Limpeza

Após a extração dos dados, foi necessário prepará-los adequadamente para a análise. Nessa etapa, foram realizadas a limpeza e o tratamento dos dados, eliminando registros incompletos, corrigindo erros, lidando com valores ausentes e tratando possíveis outliers. Além disso, foram aplicadas técnicas de transformação, como codificação de variáveis, a fim de garantir que os dados estejam em um formato adequado para a modelagem. Essa etapa abordou tratamento apenas da base de exportações, utilizada na modelagem.

Após o processo de importação e incorporação da base a um *dataframe*, foi necessário trabalhar a codificação das variáveis, tornando seus nomes mais amigáveis. Isso é importante, pois a maioria das funções de manipulação de dados utilizam-se do nome da variável como parâmetro. Para isso foi utilizado a função *rename()* presente na biblioteca *tidyverse*.

Para a etapa de modelagem de dados, tendo em vista que a sazonalidade dos dados é mensal, foi necessário agrupar as variáveis renomeadas para Ano e Mês, formando uma única denominada Data. Dessa forma, foi possível agrupar também os valores de importação da variável Valor_FOB pela referência mensal exata, somando os mesmos. Esse último agrupamento foi necessário pois, para uma mesma NCM, a base trazia mais de uma entrada por mês.

O processo de limpeza ou tratamento de dados ausentes não foi necessário, uma vez que a base não apresentava esse tipo de ocorrência. Valores considerados fora da curva (“*outliers*”) e que podem comprometer a modelagem de dados não necessariamente precisam ser tratados ao utilizar a

biblioteca *Prophet*, pois a mesma possui uma sensibilidade menor a ocorrências desse tipo de entrada. No presente trabalho esses pontos também não foram tratados.

3.5 Modelagem

Com os dados preparados, foi dado início à construção do modelo. Nessa etapa, foram aplicadas técnicas de algoritmos de aprendizado de máquina, de acordo com os objetivos do estudo.

Ao utilizar o *Prophet* em uma análise, é importante fornecer os parâmetros mínimos necessários para ajustar o modelo. Esses parâmetros garantem que a análise seja realizada corretamente e que os resultados sejam confiáveis.

O primeiro parâmetro obrigatório é o "data", que representa um dataframe contendo as observações da série temporal. Esse dataframe deve conter pelo menos duas variáveis essenciais: "ds", que indica as datas das observações, e "y", que contém os valores correspondentes da série temporal. Foram utilizadas as variáveis Data, para compor a entrada "ds", e a Valor_FOB para compor a entrada "y". Para fazer análise de apenas um único material, foi escolhido a NCM de número 73181600 que refere-se os seguintes materiais exportados: Porcas e arruelas, de ferro fundido, ferro ou aço, exceto de aço inoxidável, de diâmetro exterior inferior ou igual a 12 mm.

Além disso, existem outros parâmetros opcionais que podem ser especificados para personalizar o modelo de previsão. Por exemplo, o parâmetro "growth" define a abordagem para modelar a tendência, oferecendo opções como "linear" ou "logistic". Já o parâmetro "holidays" permite incluir informações sobre feriados relevantes para a série temporal, sendo necessário fornecer um dataframe com as datas e os nomes dos feriados. Ambos parâmetros foram ignorados nessa configuração, sendo que o primeiro parâmetro "growth" manteve-se seu valor padrão linear ao não ser configurado, e o parâmetro "holidays" não sendo necessário, pois trata-se de dados com sazonalidade mensal.

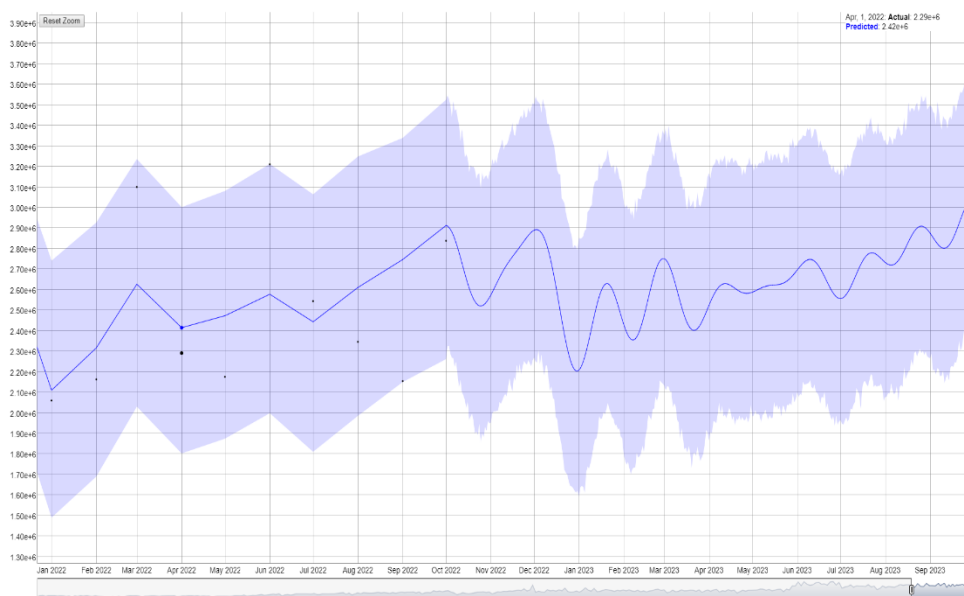
Outro parâmetro importante é o "seasonality.mode", que define a abordagem para modelar a sazonalidade, podendo ser "additive" (sazonalidade aditiva) ou "multiplicative" (sazonalidade multiplicativa). A escolha dessa

abordagem depende do tipo de série temporal e das características dos dados. Como na base utilizada a sazonalidade, nesse caso mensal, está diretamente correlacionada com a magnitude da série temporal apresentada nos dados, o parâmetro foi configurado como “multiplicative”. Portanto a função de ajuste do modelo foi: *Prophet*(NCM73181600,seasonality.mode = 'multiplicative').

Com a função ajustada foi utilizada uma outra do mesmo pacote, *make_future_dataframe()*, para previsões futuras. Essa função retorna um dataframe que contém as datas futuras com base nas datas já existentes no dataframe de treinamento e no número de períodos especificados. Ela é útil para criar um conjunto de datas futuras para o qual o modelo *Prophet* fará previsões. Por padrão é passado dois parâmetros. O primeiro dele é denominado “m” e refere-se ao modelo anteriormente ajustado. É através desse que o algoritmo irá utilizar as datas existentes no dataframe de treinamento para construir o dataframe de previsão. Um segundo parâmetro, denominado “periods”, é opcional e define a quantidade de períodos futuros que desejamos prever. Para esse caso, periods = 365 significa que está sendo solicitado previsões para os próximos 365 períodos, representando dias.

Por fim, foi utilizado a função *predict()* para as previsões conforme o período solicitado. Para ela, é passado dois parâmetros fundamentais, o primeiro, denominado “m”, contém o modelo previamente ajustado, e o segundo denominado “future”, contendo o dataframe gerado pela função anterior. Como saída, essa função gera também um dataframe de previsões resultante que possui três colunas importantes: “ds”, que contém as datas futuras; “yhat”, que representa a previsão pontual para cada data; e “yhat_lower” e “yhat_upper”, que indicam os limites inferior e superior das estimativas, fornecendo um intervalo de confiança para as previsões. Na Figura 2, o desenho da função no gráfico representa os valores de “yhat” em relação a data futura “ds” e o sombreado é dado pelo intervalo de segurança entre “yhat_lower” e “yhat_upper”.

Figura 3: Forecast para 2023



Fonte: Elaborado pelos autores

As previsões geradas podem ser utilizadas para embasar decisões estratégicas, planejamento de recursos ou até mesmo para alimentar outros modelos e análises em projetos de ciência de dados.

3.6 Análise e discussão dos resultados

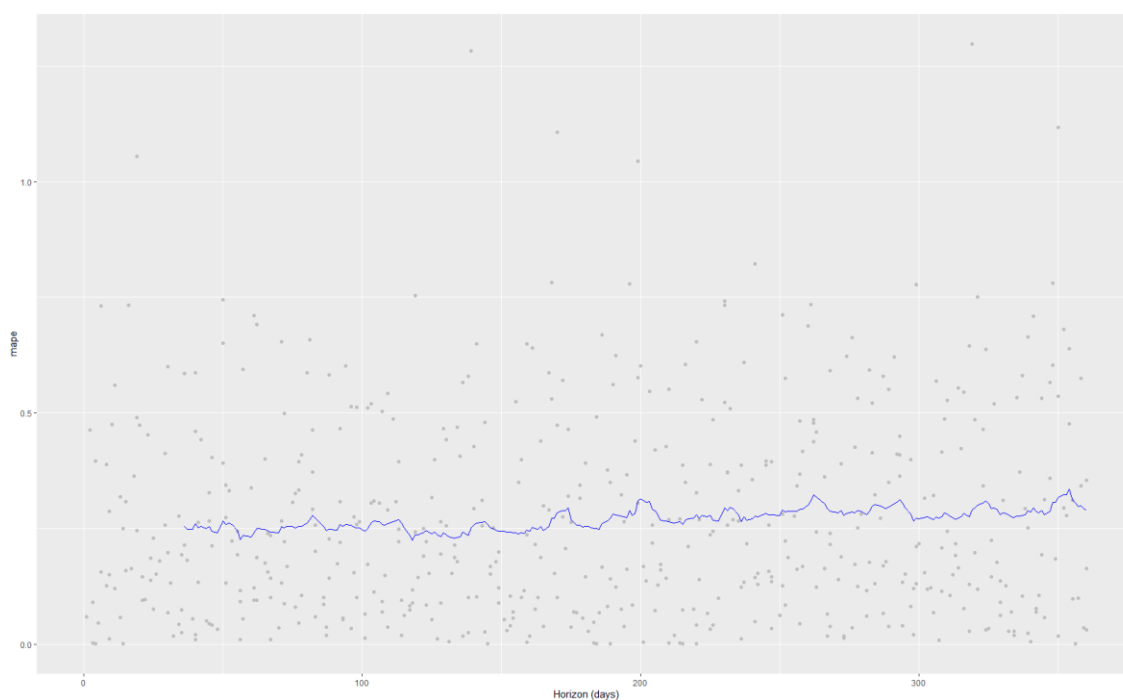
Na análise dos resultados apresentados e da eficiência do modelo para generalização de novos dados, foi utilizado processo de validação cruzada por meio da função de `cross_validation()` presente também na biblioteca do *Prophet*.

Ao utilizar a função `cross_validation()`, é necessário fornecer alguns parâmetros essenciais. O primeiro deles é o modelo *Prophet* ajustado (*m*), que contém informações importantes sobre os padrões temporais, como tendência e sazonalidade, aprendidos a partir dos dados de treinamento. Em seguida, é preciso especificar o horizonte de previsão (*horizon*), que representa o período para o qual desejamos fazer as previsões. Além disso, a função permite configurar outros parâmetros, como *initial* e *period*, que controlam a janela inicial de treinamento e o espaçamento entre os pontos de dados utilizados durante a validação cruzada.

A função `cross_validation()` retorna um dataframe que contém métricas de avaliação do modelo para cada ponto de tempo no horizonte de previsão. Essas métricas incluem o valor real da série (y), a previsão pontual do modelo ($\hat{y}$), os limites inferiores e superiores do intervalo de confiança das previsões ($\hat{y}_{lower}$ e $\hat{y}_{upper}$), e a diferença absoluta entre o valor real e a previsão ($\hat{y} - y$).

As métricas resultantes podem ser utilizadas para calcular o erro médio quadrático (MSE), o erro médio absoluto (MAE), a porcentagem média de erro absoluto entre as previsões e o valor real (MAPE) e outras métricas de avaliação. Para esse trabalho utilizou-se a função `performance_metrics()` da biblioteca do *Prophet* para cálculo de métricas MAPE. Na análise, o resultado foi plotado em um gráfico contendo uma função que expressa os valores apresentados (Figura 3).

Figura 4: MAPE



Fonte: Elaborado pelos autores

É possível observar no gráfico uma flutuação nos valores apresentados em torno de 0,25. Isso indica que 25% dos resultados apresentados podem não conferir uma previsão precisa. Essa é uma indicativa de que o modelo pode

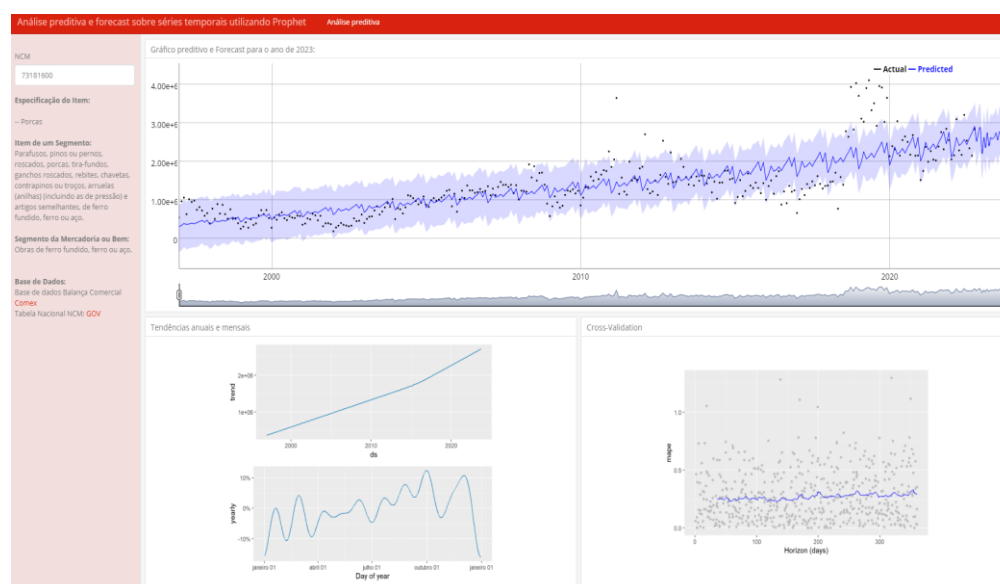
precisar de alguns ajustes, porém tratando-se de séries temporais e para a análise de como os valores de exportação vem se comportando, o modelo pode indicar tendências de comportamento em relação aos meses.

3.7 Apresentação

Os dados foram apresentados em forma de *dashboard* utilizando-se ferramentas da biblioteca *flexdashboard*. O arquivo de saída é composto por um HTML, utilizando-se o modelo *rmarkdown*.

No *dashboard* apresentado os dados apresentados em forma de gráfico, onde é possível visualizar os principais componentes, como o gráfico de preditivo e forecast para o ano de 2023. Ainda apresenta um gráfico com a tendencia de valores anuais e mensais da exportação do material selecionado. Apresenta ainda um último gráfico com a validação cruzada dos resultados, trazendo os valores da métrica MAPE, calculado. No lado esquerdo do *dashboard* é possível observar a numeração da NCM selecionada para análise, como também a divisão da descrição da mesma por especificação, segmento e item. Abaixo a Figura 4 apresenta os itens do *dashboard* gerado:

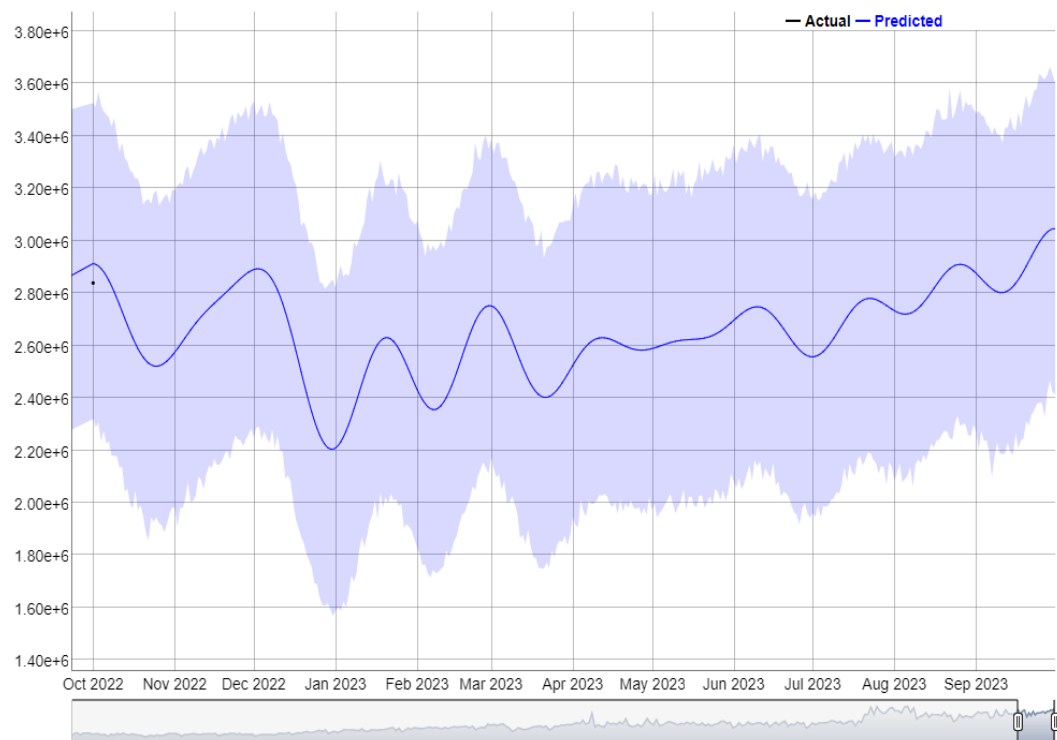
Figura 5: Dashboard



Fonte: Elaborado pelos autores.

A função `dylplot.prophet()` de plotagem do gráfico de previsão possui uma funcionalidade adicional de segmentação, permitindo visualizar exclusivamente os valores previstos para o ano de 2023. Essa característica possibilita uma análise mais precisa das tendências de comportamento dos valores específicos para esse período (Figura 5).

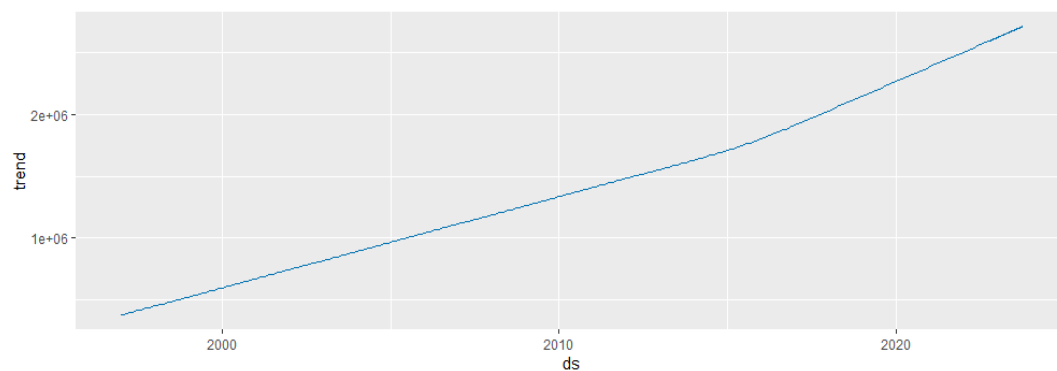
Figura 6: Visualização do comportamento dos dados preditos para 2023.



Fonte: Elaborado pelos autores.

Por fim, ainda é possível entender o comportamento dos valores no decorrer da série temporal (Figura 6) e entender as expectativas a longo prazo (anualmente), através da função `prophet_plot_components()`.

Figura 7: Tendência anual dos valores apresentados.



Fonte: Elaborado pelos autores.

4 CONCLUSÃO

A biblioteca *Prophet* é uma ferramenta poderosa e versátil para desenvolvimento de modelos de análise preditiva e previsões. Sua ampla gama de funções permite ajustar o modelo de forma precisa e realizar modelagens preditivas com facilidade. Além disso, a biblioteca oferece ferramentas de validação que auxiliam na avaliação da capacidade de generalização do modelo.

Ao combinar o *Prophet* com outras bibliotecas importantes, como o *tidyverse* para o tratamento de dados e o *flexdashboard* para visualização dos resultados, obtemos um conjunto abrangente de ferramentas de código aberto que suportam todo o processo de desenvolvimento de modelos.

O *Prophet* se destaca por sua capacidade de lidar com características comuns em séries temporais de negócios, como tendências, sazonalidades e feriados. Com uma interface intuitiva e ajuste de parâmetros flexível, o modelo pode ser adaptado às necessidades específicas de cada projeto.

A biblioteca também oferece recursos de validação cruzada que permitem avaliar o desempenho do modelo em dados não vistos. Isso proporciona uma avaliação realista da capacidade de previsão e ajuda a identificar possíveis melhorias no modelo. Para o estudo o valor apresentado nas métricas de validação pode indicar um necessário ajuste no modelo, porém o mesmo, já torna a visualização das tendências úteis para análise.

Ao utilizar o *Prophet* em conjunto com outras ferramentas de análise de dados e visualização, foi possível observar as tendências nos processos de exportação brasileira e assim tornar mais assertiva as decisões da esfera gerencial do negócio. Sua capacidade de criar modelos que prevê as flutuações nos números de exportação confere uma boa opção de desenvolvimento.

Portanto, o *Prophet*, em conjunto com outras bibliotecas de código aberto, oferece uma poderosa solução para a análise preditiva e previsão de séries temporais em diversas áreas. Sua facilidade de uso, recursos avançados e validação de desempenho tornam-no uma escolha atraente para profissionais e pesquisadores que desejam obter resultados confiáveis e embasados em suas análises.

Sugere-se, como uma possível continuação do estudo utilizando as ferramentas mencionadas, explorar sua funcionalidade de maneira a tornar o desenvolvimento mais amigável ao usuário. Além disso, seria interessante adotar uma abordagem em que as interconexões entre os gráficos do painel interajam de forma mais fluida. Para alcançar esse objetivo, poderá ser necessário estudar e incorporar novas ferramentas e bibliotecas disponíveis no ambiente R, a fim de fornecer maior autonomia e fluidez na utilização desses painéis interativos. Essa abordagem permitiria que os usuários obtivessem uma experiência mais intuitiva e eficiente ao explorar os dados e extrair insights relevantes a partir das visualizações apresentadas.

REFERÊNCIAS

BOX, G. E. P., JENKINS, G. M. Time Series Analysis: Forecasting and Control. New York: Holden Day, 1976.

FÁVERO, Luiz Paulo Lopes e BELFIORE, Patrícia Prado. Manual de análise de dados: estatística e modelagem multivariada com excel, SPSS e stata. . Rio de Janeiro: Elsevier. 2017.

GANDRUD, C. Reproducible Research with R and R Studio. 2. ed. London: CRC The R Series, 2013.

GENTLEMAN, R.; IHAKA, R.; BATES, D. The R project for statistical computing. Disponível em: . Acesso em: 10 jun. 2023

HAUSMANN, R.; PANIZZA, U.; STEIN, E. Why do countries float the way they float? Journal of Development Economics, v. 66, n. 2, 2001

Organização Mundial das Aduanas (OMA). Harmonized System (HS). Disponível em: <<https://www.wcoomd>> (Acesso em 08 Junho, 2023)

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Disponível em: <<https://www.R-project.org>> (Acesso em 10 Junho, 2023).

R PROJECT. The R Project for Statistical Computing. Disponível em: <<https://www.r-project.org/>> (Acesso em 03 Junho, 2023).

RECEITA FEDERAL. NCM, 2019. Disponível em: <<https://www.gov.br/receitafederal/pt-br/assuntos/aduana-e-comercio-exterior/classificacao-fiscal-de-mercadorias/ncm>> (Acesso em 01 Junho, 2023)

SHUMWAY, R. H., STOFFER, D. S. Time Series Analysis and Its Applications. New York: Springer, 2nd Edition, 2000.

Smith, C.W. and Stulz, R.M. (1985) The Determinants of Firms' Hedging Policies. Journal of Financial and Quantitative Analysis, 20, 391-405.

SOARES, R. O. Inflação, exportação e crescimento econômico: as experiências do México e Brasil durante o regime de câmbio flutuante. 2008. Dissertação (Mestrado) - Universidade Federal de Pernambuco, Recife, 2008.

SOUZA, R. C., CAMARGO, M. E. Análise e Previsão de Séries Temporais: Os Modelos ARIMA. Gráfica e Editora Regional, 2ª ed., Rio de Janeiro, 2004.

Taylor, S.J. (2018). *Forecasting at Scale*. The American Statistician, 72(1), 37-45.

TIPPMANN, S. Programing Tools: Adventures with R. Nature, v. 157, p. 109–110, 2015