



**Universidade Federal Fluminense**  
**Instituto de Ciência e Tecnologia**  
**Departamento de Engenharia**  
**Curso de Graduação em Engenharia de Produção**

Jean Nery dos Santos Oliveira Silva  
Pedro Costa Ceciliano

**ANÁLISE PREDITIVA DA SEGROB NOTLAD**

RIO DAS OSTRAS - RJ

2025

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1. Metodologia.....</b>                      | <b>3</b>  |
| <b>2. Estudo de Caso.....</b>                   | <b>6</b>  |
| 2.1 Entendimento do Negócio.....                | 6         |
| 2.1.1 Objetivo Geral do Negócio.....            | 6         |
| 2.1.2 Contexto.....                             | 7         |
| 2.1.3 Definindo o problema em uma pergunta..... | 7         |
| 2.1.4 5W2H.....                                 | 8         |
| 2.2 Entendimento dos Dados.....                 | 8         |
| 2.3 Preparação dos Dados.....                   | 10        |
| 2.4 Modelagem dos Dados.....                    | 11        |
| 2.4.1 Modelo Naive.....                         | 11        |
| 2.4.2 Modelo Cumulativo.....                    | 12        |
| 2.4.3 Modelo de Média Móvel.....                | 13        |
| 2.4.4 Modelo de Suavização Exponencial.....     | 14        |
| <b>3. Fundamentação Teórica.....</b>            | <b>15</b> |
| <b>4. Referências.....</b>                      | <b>15</b> |



## 1. Metodologia

A metodologia adotada para esse projeto segue os fundamentos do Cross Industry Standard Process for Data Mining (CRISP-DM). O CRISP-DM é uma abordagem estruturada e cíclica composta por seis fases principais, que garantem a organização, rastreabilidade e eficiência na execução de projetos de ciência de dados.

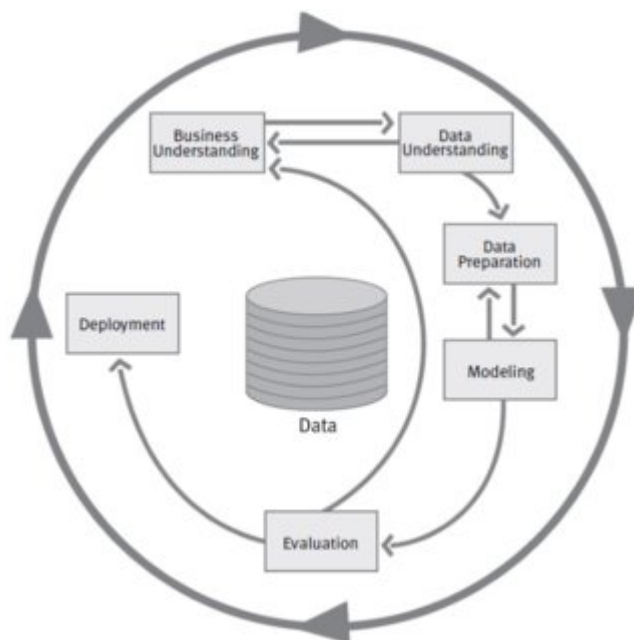


Figura 1: Fases do Modelo CRISP-DM

A primeira fase é o **Entendimento do Negócio**. Nesta fase começa com uma compreensão profunda das necessidades do cliente, incluindo seus objetivos e requisitos do projeto. Essas são algumas tarefas a serem seguidas para essa primeira fase.

- Determinar os objetivos do negócio: Entender completamente de uma visão empresarial, o que o cliente deseja realizar e em seguida definir os critérios de sucesso do negócio.
- Avaliar a situação: Determinar a disponibilidade de recursos e requisitos do projeto, avaliar os riscos e contingências e conduzir uma análise de custo-benefício.
- Determinar metas de mineração de dados: Além de definir os objetivos de negócio, deve-se também definir o que significa sucesso em uma visão técnica de mineração de dados

- Produzir plano de projeto: Selecionar tecnologias e ferramentas e definir planos detalhados para cada fase do projeto.

Por ser a primeira fase do projeto, estabelecer uma forte compreensão do negócio vai servir como fundamentação, ou seja, absolutamente essencial para o andamento do projeto.

A próxima fase é o **Entendimento dos Dados**. Somando com a primeira fase de *Entendimento do Negócio*, essa segunda fase vai direcionar o foco para identificar, coletar e analisar os conjuntos de dados que podem ajudar a atingir o objetivo do projeto. Suas tarefas são:

- Coletar dados iniciais: Adquirir os dados necessários e (se necessário) inseri-los na ferramenta de análise.
- Descrever os dados: Examinar os dados e documentar suas prioridades de superfície, como formato dos dados, números de registros ou identidades de campo.
- Explorar os dados: Se aprofundar nos dados. Consultar, visualizar e identificar as correlações entre eles.
- Verificar a quantidade dos dados: Qual nível de veracidade dos dados? Documentar qualquer problema de qualidade dos mesmos.

A terceira fase é a **Preparação de Dados**. Normalmente essa fase toma cerca de 70%, ou até mesmo 90% do tempo do projeto. É chamada por alguns de “manipulação de dados”, ela serve para preparar os conjuntos de dados finais para modelagem. Ela consiste em cinco tarefas:

- Selecionar dados: Determinar quais conjuntos de dados serão usados e documentar os motivos para inclusão/exclusão.
- Limpar dados: É uma das tarefas mais demoradas. Ela tem o objetivo de corrigir, imputar ou remover valores incorretos.
- Construir dados: Derivar novos atributos que serão úteis. Por exemplo, derivar o índice de massa corporal de alguém a partir dos campos de altura e peso.
- Integrar dados: Criar novos conjuntos de dados combinando dados de várias fontes.

- Formatar dados: Reformatar os dados conforme o necessário. Exemplo: converter valores de string que armazena números em valores numéricos para poder realizar operações matemáticas.

Depois da *Preparação de Dados*, seguimos para a **Modelagem**. Nessa fase diversos modelos serão construídos e avaliados com base em técnicas de modelagem. Esta fase tem quatro tarefas:

- Selecionar técnicas de modelagem: Determinar quais algoritmos utilizar. (por exemplo: regressão).
- Gerar design de Teste: Dependendo da modelagem, talvez seja necessário dividir os dados em conjuntos de treinamento, teste de validação.
- Construir modelo: Construir um código de programação.
- Avaliar o modelo: Interpretar os resultados do modelo com base no conhecimento do domínio, nos critérios de sucesso predefinidos e no design do teste.

O Crisp-DM sugere iterar a construção e a avaliação dos modelos até que se obtenha os melhores modelos. Porém, na prática as equipes continuam iterando até encontrar um modelo “bom o suficiente”, prosseguir pelo ciclo de vida do CRISP-DM e, então, melhorar ainda mais os modelos em iterações futuras.

A fase de **Avaliação** analisa de forma mais ampla qual modelo melhor atende aos requisitos do negócio e o que fazer em seguida. Esta fase tem três tarefas:

- Avaliar os resultados: Os modelos atendem aos critérios de sucesso do negócio? Quais devemos aprovar para o negócio?
- Processo de revisão: Revisar todo o trabalho realizado. Algo foi esquecido? Todas as etapas foram executadas corretamente? Resumir as descobertas e corrigir o que for necessário.
- Determinar as próximas etapas: Determinar se deseja prosseguir com a implantação, iterar mais, ou iniciar novos projetos, com base nas duas tarefas anteriores.

“Dependendo dos requisitos, a fase de **Implantação** pode ser tão simples quanto gerar um relatório ou tão complexa quanto implementar um processo de mineração de dados repetível em toda a empresa.” - Guia CRISP-DM. Um modelo não é particularmente útil a menos que o cliente possa acessar seus resultados.

- Planejar a implantação: Desenvolver e documentar um plano para a implantação do modelo.
- Planejar monitoramento e manutenção: Desenvolver um plano para o monitoramento e manutenção para evitar problemas durante a fase operacional (ou fase pós-projeto) de um modelo.
- Produzir relatório final: A equipe do projeto documenta um resumo do projeto, que pode incluir uma apresentação final dos resultados da mineração de dados.
- Revisar projeto: Realizar uma retrospectiva do projeto sobre o que deu certo, o que poderia ter sido melhor e como melhorar no futuro.

O Trabalho talvez não acabe depois dessas seis tarefas. Como estrutura de projeto, o CRISP-DM não define o que fazer após o projeto (também conhecido como operações).

## **2. Estudo de Caso**

### **2.1 Entendimento do Negócio**

#### **2.1.1 Objetivo Geral do Negócio**

A Segrob Notlad, uma das maiores marcas brasileiras de fast fashion, está passando por uma fase estratégica de transformação digital. Um dos pilares dessa nova fase é o uso intensivo de inteligência artificial e análise preditiva para melhorar a eficiência operacional e a assertividade nas decisões de negócio. A Segrob Notlad precisa prever o volume diário de vendas de camisetas no mês de dezembro de 2024, com base no histórico de vendas de janeiro de 2022 a novembro de 2024. Isso permitirá à empresa planejar melhor para um período de alta demanda, evitando falta de estoque ou excesso de produtos.

### 2.1.2 Contexto

A marca possui mais de 80 lojas no Brasil e presença internacional, com atuação dinâmica no mercado da moda urbana. A gestão de estoque e o planejamento de suprimentos são áreas críticas para a empresa, especialmente em períodos de alta sazonalidade, como o fim de ano. A camiseta básica é um item-chave, e falhas na previsão de demanda podem gerar rupturas de estoque ou sobras onerosas.

### 2.1.3 Definindo o problema em uma pergunta

Qual será o volume de vendas diárias de camisetas básicas durante o mês de dezembro de 2024, considerando o comportamento histórico da demanda?

Objetivos Específicos do Projeto de Mineração de Dados:

- Desenvolver um modelo preditivo capaz de estimar com precisão as vendas diárias de camisetas básicas em dezembro de 2024.
- Utilizar dados históricos de vendas (jan/2022 a nov/2024) como base de análise.
- Aumentar a agilidade e a assertividade na tomada de decisão da área de abastecimento e cadeia de suprimentos.
- Servir como um piloto para o uso sistemático de IA na previsão de demanda.

Critérios de Sucesso do Projeto:

- Capacidade do modelo de se adaptar a novas variáveis ou mudanças no cenário de negócio.
- Clareza na comunicação dos resultados para as partes interessadas.
- Potencial de escalabilidade da solução para outros produtos da marca.

Restrições e Riscos:

- Mudanças no escopo do desafio ao longo do tempo (novas variáveis, mudanças na estratégia).
- Qualidade e consistência dos dados históricos.
- Sazonalidade e eventos promocionais (Black Friday, Natal) que podem distorcer padrões históricos.
- Alinhamento entre as equipes de dados e as áreas de negócio.



Recursos e Stakeholders

- Time de análise
- Base de dados de vendas fornecida pela empresa.
- Equipes internas da Segrob Notlad nas áreas de suprimentos, marketing e estratégia digital.

2.1.4 5W2H

| Elemento           | Resposta  |
|--------------------|---|
| What (O que?)      | Prever a demanda diária de camisetas básicas.                                     |
| Why (Por quê?)     | Para otimizar o abastecimento, evitar perdas e melhorar decisões estratégicas.    |
| Who (Quem?)        | Time de análise   |
| Where (Onde?)      | Todas as lojas que comercializam a camiseta básica, com foco no Brasil.           |
| When (Quando?)     | Para o mês de <b>dezembro de 2024</b> , com base em dados de jan/2022 a nov/2024. |
| How (Como?)        | Através de modelagem preditiva utilizando técnicas de ciência de dados.           |
| How much (Quanto?) | O valor estimado da demanda diária em unidades por dia.                           |

Tabela 1 : 5W2H

Fonte: Elaboração própria.

2.2 Entendimento dos Dados

A base de dados fornecida pela empresa apresenta o volume de vendas diários para camisetas básicas masculinas entre o período de janeiro de 2022 e novembro de 2024. Com

isso, é possível analisar o comportamento desse volume de vendas ao longo de cada mês, como mostra o gráfico abaixo.

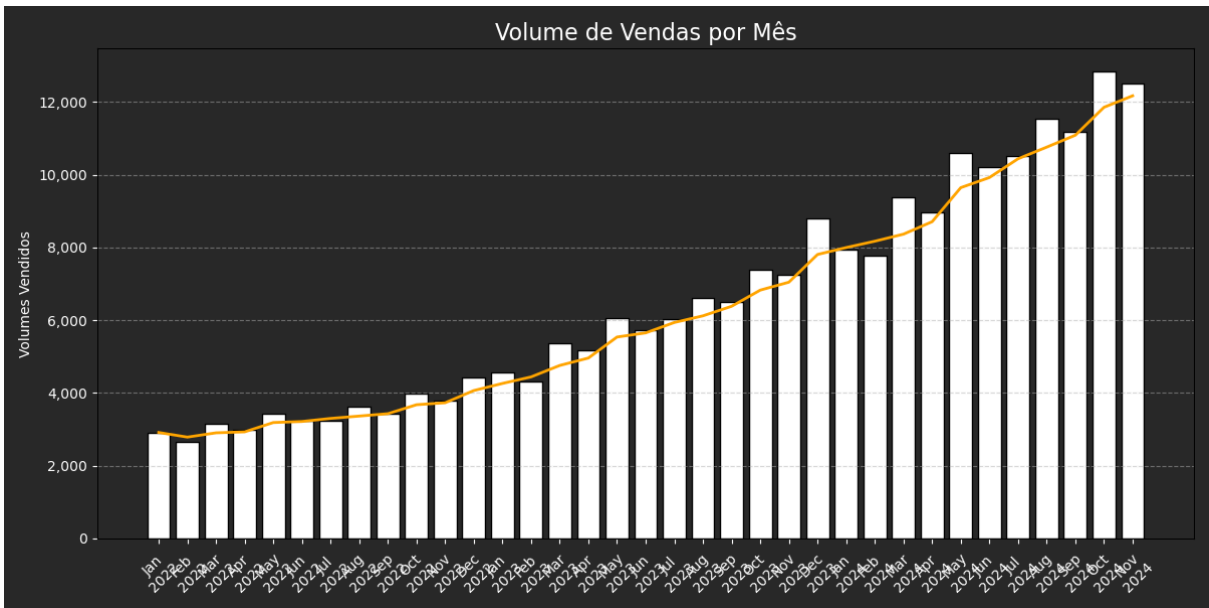


Figura 2: Transformação dos dados fornecidos em gráfico.

### Decomposição da Série Temporal - Vendas Diárias de Camisetas

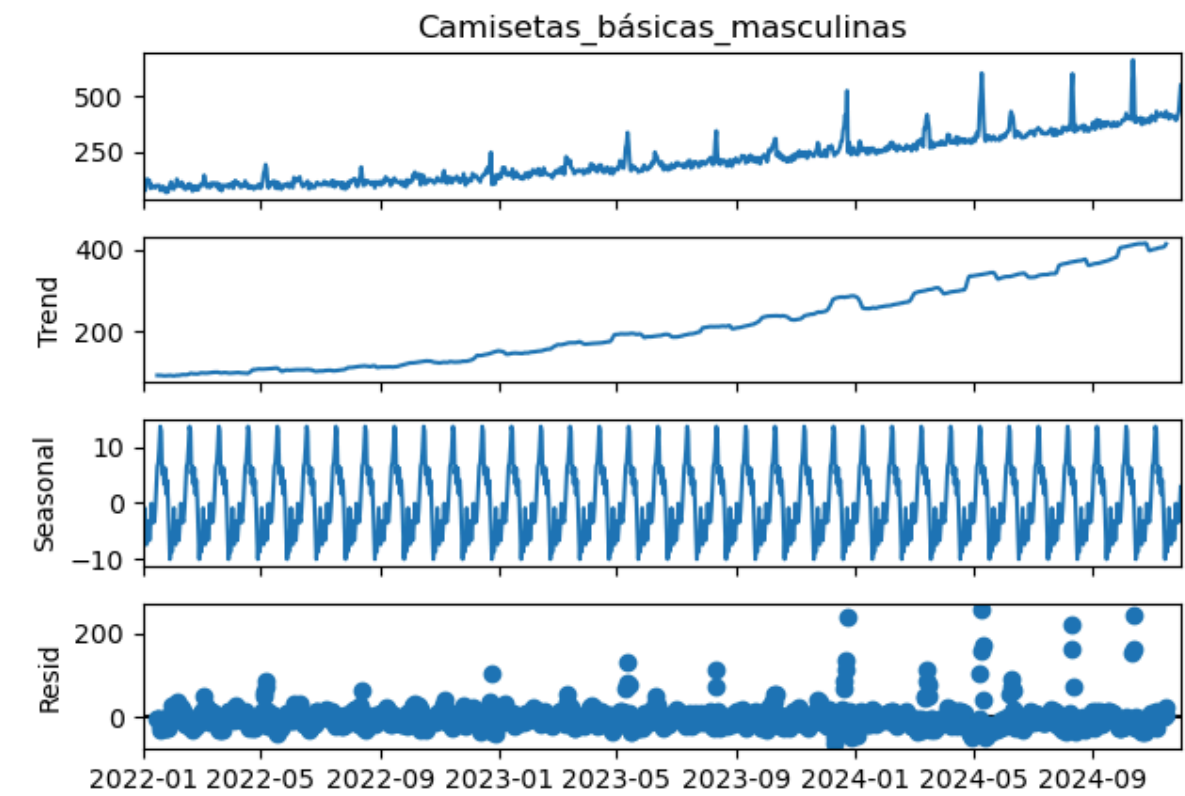


Figura 3: Gráficos para análise de tendência, sazonalidade e nível.

Analisando os gráficos observa-se um comportamento sazonal anual bem definido onde ocorre um aumento das vendas nos meses de novembro e dezembro. Isso pode estar relacionado a datas promocionais em novembro, como a Black Friday, e datas comemorativas em dezembro, como Natal, Ano Novo e recebimento do 13º salário.

Também há indícios de sazonalidade semanal em que há um maior volume de vendas nos finais de semana (sexta e sábado) e uma queda aos domingos e segundas-feiras.

Em relação a linha de tendência, constata-se uma tendência de crescimento gradual ao longo do tempo, possivelmente ligada à expansão da marca.

## 2.3 Preparação dos Dados

Antes que qualquer modelo de análise ou predição seja aplicado, é necessário que os dados estejam devidamente preparados. A qualidade e a consistência dos dados influenciam diretamente os resultados obtidos, em que qualquer discrepância em relação aos dados originais pode comprometer toda a análise. O processo de preparação de dados inclui limpeza, padronização e verificação de integridade para evitar erros ao longo do projeto e assegurar a confiabilidade das conclusões. Esse processo está alinhado com a engenharia de recursos que o faz acontecer por meio do *machine learning*.

Para dar início à preparação dos dados, foi realizada uma verificação inicial para assegurar que todas as colunas estavam devidamente preenchidas e que não havia falta de informações em nenhuma linha da base de dados original, armazenada em formato Excel. Em seguida, foi feita a checagem para identificar e remover possíveis linhas duplicadas, garantindo assim a integridade e a consistência dos dados analisados.

### PREPARAÇÃO DOS DADOS

```
from pandas import read_excel

# Carregando a planilha
path2 = r"C:\Users\Jean Nery\OneDrive\Documentos\Faculdade\9º PERÍODO\ANÁLISE PREDITIVA\25.04.22.Dados.xlsx"
df2 = read_excel(path2, sheet_name="2025.04.22")

# 1. Verificar se existem linhas com valores faltantes
linhas_faltantes = df2[df2.isnull().any(axis=1)]

if linhas_faltantes.empty:
    print("✅ Todas as linhas estão completamente preenchidas.")
else:
    print("⚠️ Existem linhas com valores ausentes:")
    print(linhas_faltantes)

# 2. Verificar duplicatas na coluna de data
coluna_data = 'Timestamp'
duplicatas_data = df2[df2.duplicated(subset=coluna_data, keep=False)]

if duplicatas_data.empty:
    print("✅ Não há valores repetidos na coluna de data.")
else:
    print("⚠️ Há valores repetidos na coluna de data:")
    print(duplicatas_data)

✅ Todas as linhas estão completamente preenchidas.
✅ Não há valores repetidos na coluna de data.
```

Figura 4: Utilização de um código em python para fazer a verificação de valores inválidos.

2.4 Modelagem dos Dados

Nesta etapa, os dados preparados na fase anterior são analisados com o objetivo de desenvolver modelos preditivos. Essa análise envolve a seleção de algoritmos apropriados, o treinamento dos modelos e a avaliação de seu desempenho, visando à obtenção de resultados confiáveis e relevantes para os objetivos do estudo.

Os modelos foram desenvolvidos com base nos dados coletados no período de janeiro de 2022 a outubro de 2024. Para a etapa subsequente, foi estabelecido que os dados referentes ao mês de novembro serão utilizados na realização dos testes, visando a seleção do modelo mais adequado.

2.4.1 Modelo Naive

O modelo Naive (ou ingênuo) é um dos modelos mais simples de previsão em séries temporais. Ele assume que o valor da próxima observação será igual ao valor mais recente. Ou seja, ele projeta que nada vai mudar.

| Item                | Descrição                                |
|---------------------|--|
| Modelo:             | Modelo Naive                             |
| Se baseia em:       | $x_t = x_{t-1} + e_t$                    |
| Onde:               | $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$ |
| Modelo de previsão: | $\hat{x}_{t,t+1} = x_t$                  |

Tabela 2: Descrição do modelo Naive

Fonte: Introdução a Séries Temporais, Dalton Borges

Explicando a tabela:

- O valor atual da série temporal  $x_t$  é igual ao valor do período anterior  $x_{t-1}$  mais um erro aleatório  $e_t$ . Esse erro representa a variação imprevisível entre um período e outro.
- $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$

- iid: erros são independentes e identicamente distribuídos, ou seja, um erro não influencia o outro e todos seguem a mesma distribuição.
- $\mu=0$ : o erro tem média zero, o que quer dizer que, em média, ele não puxa a série nem para cima nem para baixo.
- $\sigma^2=V[e]$ : o erro tem uma variância constante, indicando que a dispersão dos erros em torno da média é estável.
- $x^{t,t+1} = x_t$ 
  - A previsão para o próximo período (t+1) é simplesmente o valor atual ( $x_t$ ).
  - Isso significa que esperamos que o próximo valor da série seja igual ao último observado, sem tentar ajustar para tendências, sazonalidades ou padrões.

A implementação do modelo, por meio da linguagem Python, na base de dados da empresa, proporcionou os seguintes resultados para o mês de novembro:

| Previsões para Novembro/2024: |       |
|-------------------------------|-------|
|                               | Naive |
| 2024-11-01                    | 412   |
| 2024-11-02                    | 412   |
| 2024-11-03                    | 412   |
| 2024-11-04                    | 412   |
| 2024-11-05                    | 412   |
| 2024-11-06                    | 412   |
| 2024-11-07                    | 412   |
| 2024-11-08                    | 412   |
| 2024-11-09                    | 412   |
| 2024-11-10                    | 412   |
| 2024-11-11                    | 412   |
| 2024-11-12                    | 412   |
| 2024-11-13                    | 412   |
| 2024-11-14                    | 412   |
| 2024-11-15                    | 412   |
| 2024-11-16                    | 412   |
| 2024-11-17                    | 412   |
| 2024-11-18                    | 412   |
| 2024-11-19                    | 412   |
| 2024-11-20                    | 412   |
| 2024-11-21                    | 412   |
| 2024-11-22                    | 412   |
| 2024-11-23                    | 412   |
| 2024-11-24                    | 412   |
| 2024-11-25                    | 412   |
| 2024-11-26                    | 412   |
| 2024-11-27                    | 412   |
| 2024-11-28                    | 412   |
| 2024-11-29                    | 412   |
| 2024-11-30                    | 412   |

Figura 5: Previsão de vendas para novembro seguindo modelo naive.

### 2.4.2 Modelo Cumulativo

Diferente do modelo Naive, o modelo Cumulativo mostra a soma acumulada desses valores. Isso ajuda a identificar tendências de crescimento ou queda de forma mais clara, suavizando as flutuações pontuais. O modelo Cumulativo dá mais importância ao histórico de demanda do que o modelo Naive, em outras palavras, ele valoriza mais o passado.

| Item                | Descrição                                |
|---------------------|--|
| Modelo:             | Modelo Cumulativo                        |
| Se baseia em:       | $x_t = a + e_t$                          |
| Onde:               | $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$ |
| Modelo de previsão: | $\hat{x}_{t+1} = (\sum_{i=1}^t x_i)/t$   |

Tabela 3: Descrição do modelo Cumulativo

Fonte: Introdução a Séries Temporais, Dalton Borges

Explicando a tabela:

- Esse modelo assume que o valor da série temporal no tempo  $t$  é uma constante  $a$  mais um erro aleatório  $e_t$ . A constante  $a$  representa um nível médio fixo da série.
- $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$ 
  - iid: erros são independentes e identicamente distribuídos, ou seja, um erro não influencia o outro e todos seguem a mesma distribuição.
  - $\mu=0$ : o erro tem média zero, o que quer dizer que, em média, ele não puxa a série nem para cima nem para baixo.
  - $\sigma^2=V[e]$ : o erro tem uma variância constante, indicando que a dispersão dos erros em torno da média é estável.

- $\hat{x}_{t,t+1} = (\sum_{i=1}^t x_i)/t$ 
  - Essa fórmula representa a média aritmética dos valores observados até o tempo  $t$ .
  - É uma forma de suavizar a série, considerando todos os valores anteriores igualmente.
  - Serve para prever o próximo valor assumindo que a tendência média se mantém.

A implementação do modelo, por meio da linguagem Python, na base de dados da empresa, proporcionou os seguintes resultados para os mês de novembro:

| Previsões para Novembro/2024: |            |  |
|-------------------------------|------------|--|
|                               | Cumulativo |  |
| 2024-11-01                    | 208.289855 |  |
| 2024-11-02                    | 208.289855 |  |
| 2024-11-03                    | 208.289855 |  |
| 2024-11-04                    | 208.289855 |  |
| 2024-11-05                    | 208.289855 |  |
| 2024-11-06                    | 208.289855 |  |
| 2024-11-07                    | 208.289855 |  |
| 2024-11-08                    | 208.289855 |  |
| 2024-11-09                    | 208.289855 |  |
| 2024-11-10                    | 208.289855 |  |
| 2024-11-11                    | 208.289855 |  |
| 2024-11-12                    | 208.289855 |  |
| 2024-11-13                    | 208.289855 |  |
| 2024-11-14                    | 208.289855 |  |
| 2024-11-15                    | 208.289855 |  |
| 2024-11-16                    | 208.289855 |  |
| 2024-11-17                    | 208.289855 |  |
| 2024-11-18                    | 208.289855 |  |
| 2024-11-19                    | 208.289855 |  |
| 2024-11-20                    | 208.289855 |  |
| 2024-11-21                    | 208.289855 |  |
| 2024-11-22                    | 208.289855 |  |
| 2024-11-23                    | 208.289855 |  |
| 2024-11-24                    | 208.289855 |  |
| 2024-11-25                    | 208.289855 |  |
| 2024-11-26                    | 208.289855 |  |
| 2024-11-27                    | 208.289855 |  |
| 2024-11-28                    | 208.289855 |  |
| 2024-11-29                    | 208.289855 |  |
| 2024-11-30                    | 208.289855 |  |

Figura 6: Previsão de vendas para novembro e dezembro seguindo modelo cumulativo.

### 2.4.3 Modelo de Média Móvel

A Média Móvel é um modelo que procura generalizar os modelos Cumulativos e Naive e possui abordagens que se situam entre os extremos. Parecido com o modelo Cumulativo, porém ao invés de calcular a média de todas as observações passadas, a Média Móvel tira a média das  $M$  ultimas observações.

| Item                       | Descrição                                       |
|----------------------------|---|
| <b>Modelo:</b>             | Média Móvel                                     |
| <b>Se baseia em:</b>       | $x_t = a + e_t$                                 |
| <b>Onde:</b>               | $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$        |
| <b>Modelo de previsão:</b> | $\hat{x}_{t, t+1} = (\sum_{i=t-M+1}^t x_i) / M$ |

Tabela 4: Descrição do modelo Média Móvel

Fonte: Introdução a Séries Temporais, Dalton Borges

Explicando a tabela:

- $\hat{x}_{t, t+1} = (\sum_{i=t-M+1}^t x_i) / M$ 
  - A previsão para o próximo valor (t+1) é a média dos últimos M valores da série.
  - Essa média se move ao longo do tempo, descartando o dado mais antigo e incluindo o mais recente.

Esse modelo é muito útil sem séries de tendências fortes, para filtrar ruídos e captar o comportamento recente da série. Caso o valor de M escolhido for muito pequeno, o modelo responderá rapidamente a ruídos, e se for muito grande, perderá mudanças que não resistem por muito tempo. Normalmente utilizam valores práticos, que reflitam a unidade da escala temporal, como 1 semana, 4 meses, etc.

A implementação do modelo, por meio da linguagem Python, na base de dados da empresa, proporcionou os seguintes resultados para o mês de novembro:



| Previsões para Novembro/2024: |             |
|-------------------------------|-------------|
|                               | Média Móvel |
| 2024-11-01                    | 414.533333  |
| 2024-11-02                    | 415.251111  |
| 2024-11-03                    | 416.059481  |
| 2024-11-04                    | 416.761464  |
| 2024-11-05                    | 416.586846  |
| 2024-11-06                    | 417.039741  |
| 2024-11-07                    | 417.374399  |
| 2024-11-08                    | 418.020213  |
| 2024-11-09                    | 418.620886  |
| 2024-11-10                    | 419.408249  |
| 2024-11-11                    | 414.455191  |
| 2024-11-12                    | 406.237031  |
| 2024-11-13                    | 400.544932  |
| 2024-11-14                    | 401.129763  |
| 2024-11-15                    | 401.067421  |
| 2024-11-16                    | 401.636335  |
| 2024-11-17                    | 402.124213  |
| 2024-11-18                    | 402.295020  |
| 2024-11-19                    | 402.704854  |
| 2024-11-20                    | 402.695016  |
| 2024-11-21                    | 403.718183  |
| 2024-11-22                    | 404.675456  |
| 2024-11-23                    | 404.897971  |
| 2024-11-24                    | 405.661237  |
| 2024-11-25                    | 406.549945  |
| 2024-11-26                    | 407.001610  |
| 2024-11-27                    | 407.301663  |
| 2024-11-28                    | 407.645052  |
| 2024-11-29                    | 408.566554  |
| 2024-11-30                    | 409.085439  |

Figura 7: Previsão de vendas para novembro seguindo modelo de média móvel.

#### 2.4.4 Modelo de Suavização Exponencial

O modelo de Suavização Exponencial é uma técnica que prevê valores futuros com base em valores passados, dando mais peso para os dados mais recentes e menos peso para os antigos, por isso o nome “exponencial”, pois a importância dos dados antigos decai exponencialmente.

A suavização exponencial simples implica em :

- Demanda estacionária , sem tendência nem sazonalidade. Considera apenas o nível da demanda.
- O valor das observações diminui com o tempo.
- Utiliza uma constante de suavização  $\alpha$  , onde  $0 \leq \alpha \leq 1$ .
- Na pratica  $0,1 \leq \alpha \leq 0,3$ .

| Item    | Descrição                      |
|---------|--------------------------------|
| Modelo: | Suavização Exponencial Simples |

|                            |   |
|----------------------------|---|
| <b>Se baseia em:</b>       | $x_t = a + e_t$   |
| <b>Onde:</b>               | $e_t \sim iid(\mu = 0, \sigma^2 = V[e])$  |
| <b>Modelo de previsão:</b> | $\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t, \text{ com } 0 \leq \alpha \leq 1$ |

Tabela 4: Descrição do modelo Suavização Exponencial Simples

Fonte: Suavização Exponencial I, Dalton Borges

Explicando a tabela:

- $\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t, \text{ com } 0 \leq \alpha \leq 1$ 
  - $\hat{x}_{t+1}$ : previsão do próximo valor (tempo t+1) feita no tempo t.
  - $\alpha$ : fator de suavização (entre 0 e 1).
  - Se  $\alpha \approx 1$ : mais peso ao valor atual  $\rightarrow$  previsão mais reativa.
  - Se  $\alpha \approx 0$ : mais peso à previsão passada  $\rightarrow$  previsão mais suave.

A implementação do modelo, por meio da linguagem Python, na base de dados da empresa, proporcionou os seguintes resultados para o mês de novembro:

| Previsões para Novembro/2024: |                 |
|-------------------------------|-----------------|
|                               | Suavização Exp. |
| 2024-11-01                    | 398.297011      |
| 2024-11-02                    | 395.933658      |
| 2024-11-03                    | 397.588005      |
| 2024-11-04                    | 396.429962      |
| 2024-11-05                    | 397.240592      |
| 2024-11-06                    | 396.673151      |
| 2024-11-07                    | 397.070360      |
| 2024-11-08                    | 396.792314      |
| 2024-11-09                    | 396.986946      |
| 2024-11-10                    | 396.850704      |
| 2024-11-11                    | 396.946073      |
| 2024-11-12                    | 396.879315      |
| 2024-11-13                    | 396.926046      |
| 2024-11-14                    | 396.893334      |
| 2024-11-15                    | 396.916232      |
| 2024-11-16                    | 396.900203      |
| 2024-11-17                    | 396.911424      |
| 2024-11-18                    | 396.903569      |
| 2024-11-19                    | 396.909067      |
| 2024-11-20                    | 396.905219      |
| 2024-11-21                    | 396.907913      |
| 2024-11-22                    | 396.906027      |
| 2024-11-23                    | 396.907347      |
| 2024-11-24                    | 396.906423      |
| 2024-11-25                    | 396.907070      |
| 2024-11-26                    | 396.906617      |
| 2024-11-27                    | 396.906934      |
| 2024-11-28                    | 396.906712      |
| 2024-11-29                    | 396.906867      |
| 2024-11-30                    | 396.906759      |

Figura 8: Previsão de vendas para novembro seguindo modelo de suavização exponencial simples.

### 3. Fundamentação Teórica

Nessa etapa do projeto, a análise e avaliação de modelos preditivos é essencial para compreender as métricas utilizadas para mensurar a precisão das previsões, bem como os métodos de validação empregados para garantir sua confiabilidade.

A qualidade das previsões de um modelo pode ser mensurada por meio de diversas métricas estatísticas, tais como o MAPE (Mean Absolute Percentage Error – Erro Percentual Médio Absoluto), o RMSE (Root Mean Squared Error – Raiz do Erro Quadrático Médio) e o MAD (Mean Absolute Deviation – Desvio Absoluto Médio). > Para a aplicação dessas métricas, serão comparados os valores referentes ao mês de novembro, tomando como base os valores reais e os valores previstos, possibilitando uma análise precisa do desempenho do modelo.

#### 3.1 MAPE

Essa métrica faz o cálculo da média dos erros absolutos expressos em termos percentuais em relação aos valores reais. Quanto menor o MAPE, melhor a precisão do modelo.

$$MAPE = \frac{\sum(\frac{|y_i - \hat{y}_i|}{y_i})}{n}$$

- $y_i$  - valor real;
- $\hat{y}_i$  - valor previsto;
- $n$  - total de observações.

#### 3.2 RMSE

Essa métrica avalia a magnitude dos erros elevando ao quadrado as diferenças entre os valores reais e previstos.

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

### 3.3 MAD

Essa métrica mede a dispersão dos erros, calculando a média das diferenças absolutas entre os valores reais e os valores previstos.

$$MAD = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

### 3.4 Validação

A validação dos modelos de previsão é uma etapa essencial para garantir a confiabilidade das estimativas realizadas. Para isso, foram calculadas métricas estatísticas por meio da linguagem de programação Python, que avaliam o desempenho de cada abordagem utilizada para prever os valores referentes ao mês de novembro de 2024.

```
### Erros de Previsão - Novembro 2024 ###  
  
Modelo: Naive  
MAD: 15.40  
RMSE: 31.07  
MAPE (%): 3.38  
  
Modelo: Cumulativo  
MAD: 208.31  
RMSE: 210.56  
MAPE (%): 49.78  
  
Modelo: Média Móvel  
MAD: 16.71  
RMSE: 32.22  
MAPE (%): 3.68  
  
Modelo: Suavização Exp.  
MAD: 21.07  
RMSE: 36.58  
MAPE (%): 4.65
```

Figura 9: Erros das previsões dos modelos.

- O modelo Naive apresentou os menores valores de erro, indicando previsões mais próximas dos valores reais.
- O modelo Cumulativo teve os maiores erros, demonstrando baixa precisão nas previsões.
- A Média Móvel obteve resultados similares ao modelo Naive, sugerindo que pode ser uma alternativa viável.
- A Suavização Exponencial apresentou erros superiores aos modelos Naive e Média Móvel, mas ainda inferiores ao Cumulativo.

Portanto, a validação dos modelos indica que o Naive e a Média Móvel são as abordagens mais precisas para estimar os valores de novembro de 2024. Com base nesses resultados, foram escolhidos os modelos que melhor equilibram a precisão e estabilidade para a próxima etapa do processo.

| Previsões para Dezembro/2024: |       |             |
|-------------------------------|-------|-------------|
|                               | Naive | Média Móvel |
| 2024-12-01                    | 412   | 408.988287  |
| 2024-12-02                    | 412   | 408.803452  |
| 2024-12-03                    | 412   | 408.588530  |
| 2024-12-04                    | 412   | 408.339499  |
| 2024-12-05                    | 412   | 408.058766  |
| 2024-12-06                    | 412   | 407.774497  |
| 2024-12-07                    | 412   | 407.465656  |
| 2024-12-08                    | 412   | 407.135364  |
| 2024-12-09                    | 412   | 406.772536  |
| 2024-12-10                    | 412   | 406.377591  |
| 2024-12-11                    | 412   | 405.943236  |
| 2024-12-12                    | 412   | 405.659504  |
| 2024-12-13                    | 412   | 405.640253  |
| 2024-12-14                    | 412   | 405.810097  |
| 2024-12-15                    | 412   | 405.966108  |
| 2024-12-16                    | 412   | 406.129398  |
| 2024-12-17                    | 412   | 406.279166  |
| 2024-12-18                    | 412   | 406.417665  |
| 2024-12-19                    | 412   | 406.555086  |
| 2024-12-20                    | 412   | 406.683427  |
| 2024-12-21                    | 412   | 406.816374  |
| 2024-12-22                    | 412   | 406.919647  |
| 2024-12-23                    | 412   | 406.994454  |
| 2024-12-24                    | 412   | 407.064336  |
| 2024-12-25                    | 412   | 407.111106  |
| 2024-12-26                    | 412   | 407.129812  |
| 2024-12-27                    | 412   | 407.134085  |
| 2024-12-28                    | 412   | 407.128499  |
| 2024-12-29                    | 412   | 407.111281  |
| 2024-12-30                    | 412   | 407.062772  |
| 2024-12-31                    | 412   | 406.995349  |

Figura 10: Previsão de vendas para para o mês de dezembro seguindo os modelos naive e média móvel.

#### 4. Referências

- DATA SCIENCE PM. *O que é CRISP DM?* Data Science PM, 9 dez. 2024. Disponível em: <https://www.datascience-pm.com/crisp-dm-2/>. Acesso em: 18 maio 2025.
- Zheng, Alice; CASARI, Amanda. *Feature engineering for machine learning: principles and techniques for data scientists*. 1. ed. Sebastopol: O'Reilly Media, 2018.

