Séries Temporais

Definição

Uma série temporal é uma sequência de observações registradas em momentos específicos no tempo.
Exemplo: vendas diárias de um produto, temperaturas mensais ou valores de ações ao longo de semanas.

Estrutura

Observações (yt): Valores medidos em momentos t.

Tempo (t): Pode ser diário, semanal, mensal, anual ou em intervalos irregulares.

Utilidade e Quando Usar Séries Temporais

Séries temporais são úteis em cenários onde os dados variam ao longo do tempo e existe a necessidade de entender padrões, realizar análises ou prever eventos futuros. Elas são empregadas em diversas áreas devido à sua capacidade de capturar tendências, sazonalidades e dependências temporais.

A análise de séries temporais é essencial para previsão, identificação de padrões, monitoramento, tomada de decisão estratégica e otimização de processos.

A previsão permite estimar valores futuros com base em dados históricos, sendo útil para demandas de estoque, vendas, produção de energia e preços de mercado. Já a análise de padrões identifica tendências, sazonalidades e ciclos, auxiliando no entendimento de variações de vendas, padrões climáticos e comportamento do consumidor.

O monitoramento detecta anomalias em tempo real, como falhas em máquinas, fraudes financeiras e variações no tráfego de redes. Na tomada de decisão estratégica, essas informações ajudam no planejamento de campanhas, precificação dinâmica e infraestrutura urbana.

Por fim, a otimização permite ajustar processos automaticamente, como controlar a produção industrial, gerenciar o consumo de energia e otimizar transportes públicos.

Componentes de Séries Temporais

As séries temporais podem ser decompostas em quatro componentes principais:

1. Tendência (Trend)

Representa o movimento de longo prazo na série. Exemplo: Crescimento consistente no número de vendas ao longo dos anos.

2. Sazonalidade (Seasonality)

Padrões que se repetem em intervalos regulares de tempo. Exemplo: Aumento de vendas no Natal.

3. Ciclo (Cycle)

Flutuações que ocorrem devido a fatores econômicos ou outros ciclos (não tem a ver com o tempo). Exemplo: Alta e baixa de um mercado financeiro em períodos não fixos.

4. Resíduo (Noise)

Variação aleatória ou irrelevante na série. Exemplo: Pequenas flutuações inesperadas em vendas diárias.

Tipos de Séries Temporais

1. **Univariada**

   Contém apenas uma variável observada ao longo do tempo. O foco é entender e prever essa única variável com base em seus próprios valores passados. Exemplo: Preço diário de uma ação, temperatura média mensal, vendas de um produto ao longo do ano.

2. **Multivariada**

   Envolve múltiplas variáveis inter-relacionadas, analisadas simultaneamente. Essas variáveis podem influenciar umas às outras, tornando a modelagem mais complexa. Exemplo: Preço de uma ação considerando também taxas de juros e volume de negociação, previsão de demanda levando em conta fatores climáticos e promoções.

Estacionaridade

Um conjunto de dados estacionário é aquele cujas propriedades estatísticas não mudam ao longo do tempo. Isso significa que a média, a variância e a auto covariância do processo são constantes. Em outras palavras, uma série temporal estacionária não apresenta tendências ou sazonalidades de longo prazo, tornando-a mais previsível e mais fácil de modelar.

Modelos e Técnicas de Análise

Modelos Estatísticos

1. **Média Móvel (Moving Average)**

   A Média Móvel suaviza flutuações em uma série temporal, facilitando a identificação de tendências. Ela calcula a média de um número fixo de valores anteriores, eliminando variações aleatórias.

2. **ARIMA (AutoRegressive Integrated Moving Average)**

   O modelo ARIMA combina três componentes principais para modelar séries temporais:

   AR (AutoRegressive): Usa valores passados da própria série para prever futuros.

   I (Integrated): Aplica diferenciação para tornar a série estacionária.

   MA (Moving Average): Modela os erros da previsão baseando-se em valores passados.

   O ARIMA é indicado para séries temporais estacionárias ou que podem ser transformadas em estacionárias. Ele é representado como **ARIMA(p, d, q)**, onde:

   p = Número de defasagens na parte autoregressiva.

   d = Número de diferenciações para tornar a série estacionária.

   q = Número de termos de média móvel.

   Exemplo de Aplicação: Previsão de demanda por produtos ou previsão de inflação.

3. **SARIMA (Seasonal ARIMA)**

   O **SARIMA** é uma extensão do ARIMA que incorpora sazonalidade, ou seja, padrões que se repetem periodicamente. Ele adiciona quatro novos parâmetros para capturar variações sazonais:

   P = Ordem da parte autoregressiva sazonal.

   D = Diferenciação sazonal necessária.

Q = Ordem da média móvel sazonal.

S = Período da sazonalidade (exemplo: 12 para dados mensais).

O modelo é representado como **SARIMA(p, d, q) × (P, D, Q, S)**.

Exemplo de Aplicação: Previsão de vendas sazonais, como consumo de energia elétrica ao longo do ano.

Etapas Práticas para Análise de Séries Temporais

1. Importar e Visualizar os Dados

A visualização de dados é uma etapa essencial na análise de séries temporais, pois permite identificar padrões, tendências, sazonalidade e anomalias antes da modelagem. Algumas das principais técnicas incluem:

Gráfico de Linha: A forma mais comum de visualização, exibindo a evolução dos dados ao longo do tempo. Ajuda a detectar tendências e padrões sazonais.

Histogramas e Boxplots: Úteis para analisar a distribuição dos dados e identificar outliers.

Autocorrelação (ACF) e Parcial (PACF): Mostram como os valores passados influenciam os valores futuros, auxiliando na escolha de modelos ARIMA/SARIMA.

Decomposição de Série Temporal: Separa os componentes da série em tendência, sazonalidade e resíduo, facilitando a interpretação.

Heatmaps e Gráficos de Dispersão: Podem ser usados para visualizar padrões sazonais ou relações entre múltiplas variáveis em séries temporais multivariadas.

2. Analisar a Estacionaridade

Visualização Gráfica

Gráfico de Linha: Se houver uma tendência crescente/decrescente ou padrões sazonais evidentes, a série pode não ser estacionária.

Rolling Statistics: Calcular e plotar a média e a variância em janelas móveis para ver se se mantêm constantes.

Testes Estatísticos

Teste de Dickey-Fuller Aumentado (ADF)

Hipótese nula (H0): A série tem raiz unitária (não estacionária).

Hipótese alternativa (H1): A série é estacionária.

Se o p-valor for menor que 0,05, rejeitamos H0 e concluímos que a série é estacionária.

Teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin)

Hipótese nula (H0): A série é estacionária.

Hipótese alternativa (H1): A série não é estacionária.

Se o p-valor for menor que 0,05, rejeitamos H0, indicando que a série não é estacionária.

Teste de Phillips-Perron (PP)

Similar ao ADF, mas mais robusto a heterocedasticidade.

Função de Autocorrelação (ACF) e Autocorrelação Parcial (PACF)

Se a ACF decai lentamente em vez de cair rapidamente para zero, a série pode ser não estacionária.

Como Tornar uma Série Estacionária?

Se a série não for estacionária, pode-se aplicar transformações para estabilizar suas propriedades:

Diferenciação: Subtrair o valor anterior do atual ($Y_t - Y_{t-1}$) para remover tendências.

Transformação Logarítmica ($\log Y_t$): para estabilizar a variância.

Diferenciação Sazonal: Para séries com sazonalidade, subtrair valores do mesmo período anterior ($Y_t - Y_{t-S}$).

Remover Tendência com Modelos de Regressão: Ajustar e remover uma tendência linear ou polinomial.

Yt representa o valor da série temporal no instante t. Ou seja, é o valor observado no tempo t.

S representa a sazonalidade da série, ou seja, o número de períodos após os quais os padrões se repetem.

3. Decomposição

A decomposição de séries temporais é uma técnica que separa a série em diferentes componentes para facilitar a análise e a modelagem. Isso permite entender melhor as tendências, identificar padrões sazonais e remover ruídos. Podendo ser dividida em Tendencia, Sazonalidade e Resíduo.

1. Decomposição Aditiva

Quando os componentes são somados:

$$Y_t = \text{Trend}_t + \text{Seasonal}_t + \text{Residual}_t$$

Usada quando a **amplitude da sazonalidade é constante** ao longo do tempo.

Exemplo: Temperatura média diária ao longo do ano.

2. Decomposição Multiplicativa

Quando os componentes são multiplicados:

$$Y_t = \text{Trend}_t \times \text{Seasonal}_t \times \text{Residual}_t$$

Aplicada quando a **sazonalidade varia em intensidade** conforme a tendência cresce ou diminui.

Exemplo: Receita de uma empresa, onde picos sazonais aumentam à medida que a empresa cresce.

4. Modelagem

Escolher e ajustar o modelo apropriado (ARIMA, SARIMA, LSTM, Prophet, etc.).

5. Avaliação

Ao trabalhar com séries temporais, a separação dos dados deve ser feita de forma **sequencial**, mantendo a ordem cronológica. Diferente de problemas comuns de aprendizado de máquina, onde a divisão pode ser aleatória, em séries temporais isso comprometeria a capacidade preditiva do modelo.

Como Dividir os Dados?

A abordagem mais utilizada é:

- **Conjunto de Treino**: Parte inicial dos dados, usada para ajustar o modelo.
- **Conjunto de Teste**: Últimos períodos da série, utilizados para avaliar a performance.

Métricas de Avaliação para Séries Temporais

Depois de treinar o modelo, ele precisa ser avaliado com métricas que quantificam o erro das previsões em relação aos valores reais. As mais utilizadas são:

1. RMSE (Erro Quadrático Médio - Root Mean Squared Error)

Mede a diferença média entre valores reais e previstos, penalizando mais os erros grandes:

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

- Valores menores indicam previsões mais precisas.
- Penaliza fortemente erros grandes devido à elevação ao quadrado.

2. MAE (Erro Absoluto Médio - Mean Absolute Error)

Mede a média dos erros absolutos, sem dar peso maior para erros grandes:

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

- Fácil de interpretar, pois está na mesma unidade dos dados.
- Não amplifica tanto os erros maiores, diferente do RMSE.

3. MAPE (Erro Percentual Médio Absoluto - Mean Absolute Percentage Error)

Expressa o erro médio em termos percentuais, o que facilita a interpretação:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - p_i}{y_i} \right|$$

Útil para comparar previsões em diferentes escalas de valores.

Problemático quando Yi se aproxima de zero, pois pode gerar distorções.

Desafios em Séries Temporais

1. **Não-estacionaridade**

   Séries temporais que possuem média ou variância variando ao longo do tempo.

   Solução: Transformações (diferenciação, logaritmos).

2. **Dados Perdidos (Missing Data)**

   Dados ausentes podem prejudicar a análise.

   Solução: Interpolação ou preenchimento com valores estimados.

3. **Séries Curta**

   Poucos dados dificultam a identificação de padrões.

   Solução: Adicionar variáveis externas ou usar modelos que lidem bem com pequenos conjuntos de dados.

4. **Sazonalidades Complexas**

   Séries com múltiplos padrões sazonais podem ser difíceis de modelar.