

Muito além da técnica

Habilidades essenciais para estatísticos

MAE0413 - Estatística Aplicada I

Magno T. F. Severino

Agenda

- Minha trajetória
- Caso de estudo
- Habilidades Técnicas Fundamentais
- *Soft Skills* Fundamentais

Minha Trajetória: Da Universidade ao Mercado

**Conclusão Bacharelado
Matemática
Computacional**

UFMG

2015



**Finalista no Concurso de
Melhor Dissertação de
Mestrado**

SINAPE 2018

2018



**Conclusão Mestrado
em Estatística**

“Extending JAGS for spatial data”
UFMG

2018



**Professor de Pós-
Graduação Lato-Sensu**

Inspere

2021



**Especialista no Banco
BTG Pactual**

Precificação de derivativos

2022



**Conclusão Doutorado em
Probabilidade e Estatística**

“Estimation and model selection for
graphical models under mixing conditions”
IME USP

2024



Especialista na Pefisa
Ciência de dados

2024



guia da pós

IA valoriza capacidade técnica de estatísticos

Além de especializações, mestrado e doutorado viram opção para profissionais já inseridos no mercado

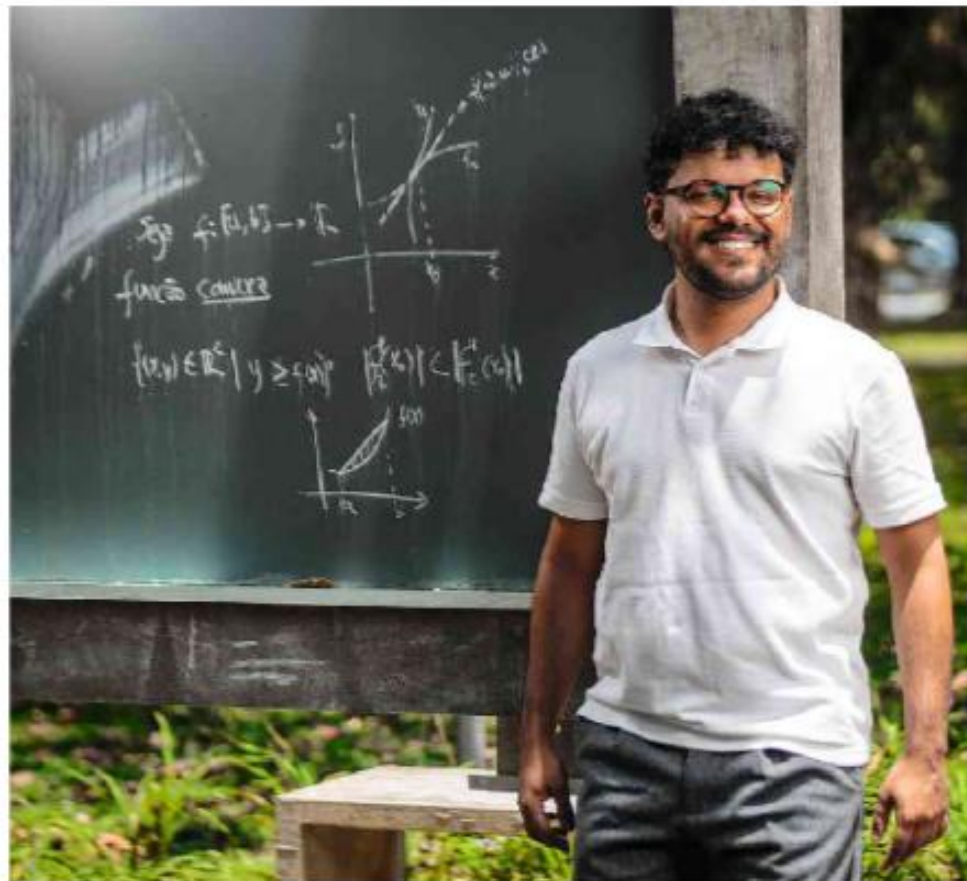
Beatriz Gatti

SÃO PAULO Diante de um volume de dados cada vez maior, empresas de variados segmentos aumentam a demanda por estatísticos com alto conhecimento técnico. Para acompanhar o movimento e se destacar no mercado, os profissionais recorrem a programas de pós-graduação, que por sua vez têm se aprofundado em conceitos de machine learning e outros aspectos relacionados à inteligência artificial.

É o caso de Magno Tairone, 31, que trabalha como analista de modelagem estatística em uma empresa do mercado financeiro, em São Paulo, e é professor de uma especialização em ciência de dados no Insper.

Natural de Belo Horizonte, Magno concluiu a graduação em matemática computacional e fez mestrado em estatística, ambos pela Universidade Federal de Minas Gerais (UFMG). Foi na pós-graduação que ele começou a estudar machine learning, técnica hoje muito utilizada em seu dia a dia na companhia.

Com uma base de dados de milhões de clientes, Magno cria mo-



Magno Tairone, mestre e doutor em estatística, trabalha como analista de modelagem estatística e professor do Programa Avançado de Data Science e Decisão do Insper. Jardiel Carvalho/Folhapress

delos preditivos relacionados ao risco de crédito, ou seja, à probabilidade de cada pessoa não pagar as faturas em dia.

"Eu observo, por exemplo, quantas vezes determinado cliente atrasou a fatura meses depois de fazer o cartão e crio uma régua a partir disso para saber se a pessoa [que solicita crédito] pode ser boa ou má pagadora", explica.

A estatística é utilizada no processamento das informações. "Eu faço a máquina aprender os padrões e a equipe de TI (tecnologia da informação) transforma o modelo em uma solução que permite esse cálculo de probabilidade em um apertar de botão", acrescenta.

Muitas vezes confundidas entre si, estatística, tecnologia da informação e ciência de dados estão profundamente correlacionadas. O profissional de TI cria ferramentas práticas a partir da base teórica desenvolvida pelos estatísticos. A ciência de dados engloba os dois campos e ainda inclui a área de negócios, que solicita as demandas por modelos ou soluções de TI.

Continua na pág. A45

Estudo de caso: segmentação de clientes

Estudo de caso: segmentação de clientes

Analise do **comportamento de compra** dos clientes que consomem produtos de vestuário e lar, categorizando-os com base em suas preferências de compra, tanto **com desconto** quanto **sem desconto**.

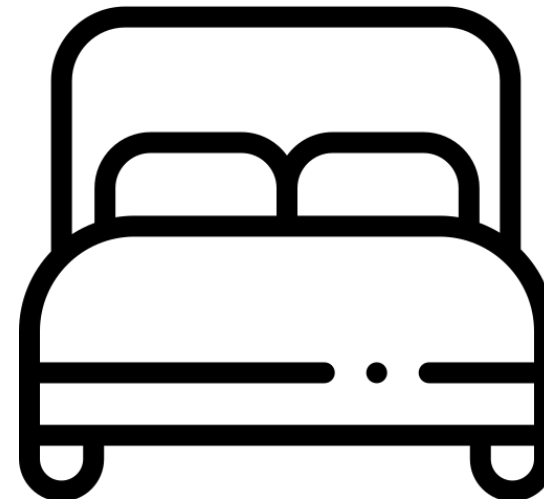
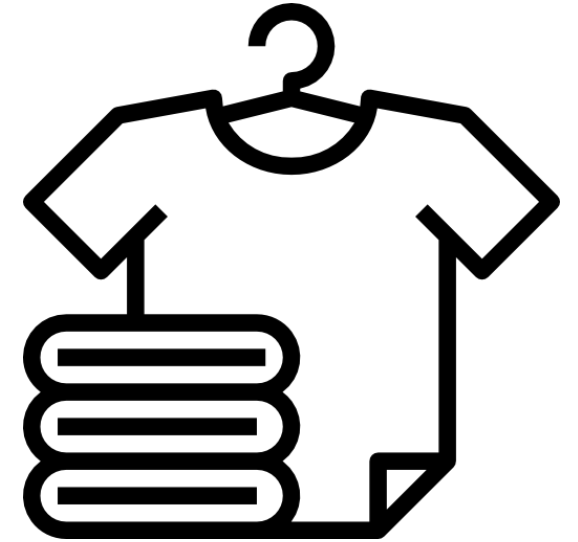
Objetivo: otimizar a oferta e disponibilização de cupons de desconto.



Estudo de caso: segmentação de clientes

Informações disponíveis:

- Número de compras nos últimos 12 meses
- Valor total de cada compra
- Quantidades de itens comprados com desconto
- Desconto aplicado em cada item
- Valor médio por compra
- Gastos em itens de Lar
- Gastos em itens de Vestuário
- Intervalo médio entre compras
- Dias desde a última compra



Habilidades Técnicas Fundamentais

Entendimento do problema

Desafio: o time de marketing queria otimizar campanhas de cupons, mas não sabia se todos os clientes tinham o mesmo comportamento.

- **Pergunta central:** Quem são os clientes que utilizam cupons? Eles são diferentes dos que não usam?
- **Hipótese inicial:** Pode haver perfis distintos de clientes, e segmentá-los ajudaria na estratégia promocional.



Antes de qualquer modelagem, é essencial entender a necessidade do negócio e formular a pergunta certa.

Coleta e Análise Exploratória dos Dados

Os dados disponíveis são suficientes para obter a solução desejada?

- Os dados de compras e cupons estavam em **diferentes bases de dados** e precisavam ser combinados.
- **SQL** para **extrair** os dados de compras e cupons de diferentes tabelas.
- **PySpark** foi usado para lidar com o grande volume de dados (~milhões de registros), garantindo eficiência no processamento.
- Técnicas de **limpeza** e **tratamento** de dados (imputação de valores, tratamento de outliers, feature engineering).
- Tratamos **valores nulos** e criamos variáveis derivadas, como "número de compras com cupom" e "ticket médio com/sem cupom".

 Um bom modelo começa com dados bem preparados. Saber manipular grandes volumes de informação é essencial para análises robustas.

Escolha da Técnica de Segmentação

- **Cluster Hierárquico:** Exploramos dendrogramas para visualizar os agrupamentos e optamos por um K-Means ajustado com base nessa análise.
- **K-Means:** Inicialmente testamos essa técnica, mas percebemos que os grupos não refletiam bem os padrões de uso dos cupons.



Escolher o modelo certo exige testes, ajustes e alinhamento com o objetivo do negócio.

Interpretação dos Clusters e Implicações no Negócio

Como transformar um resultado estatístico em uma ação concreta para o time de marketing?

Após a segmentação, identificamos **três perfis principais de clientes**:

- **Cluster 1 (Caçadores de Promoção):** Compram apenas com cupom, ticket médio baixo.
- **Cluster 2 (Clientes Oportunistas):** Compram com e sem cupom, ticket médio variável.
- **Cluster 3 (Fidelizados sem Cupom):** Compram com frequência, independente de desconto.

Impacto no negócio:


- Para o **Cluster 1**, sugerimos limitar o número de cupons para evitar prejuízos.
- Para o **Cluster 3**, criamos estratégias de fidelização sem desconto.



Modelos estatísticos só têm valor quando geram
insights acionáveis.

Produtização da solução

- **Integração com sistemas:** O modelo precisa ser implementado em uma estrutura que permita consultas rápidas e automáticas.
- Como um **novo cliente** que entra no sistema pode ser classificado corretamente?
- O comportamento dos clientes muda com o tempo → o modelo precisa ser **reavaliado** periodicamente.
- Equilibrar **complexidade do modelo** com **eficiência operacional** (modelos muito complexos podem ser difíceis de rodar em tempo real).

 Um modelo de machine learning não termina quando se encontra um bom ajuste estatístico → ele precisa ser viável operacionalmente e útil no dia a dia do negócio.

Monitoramento e Governança de Modelos

- **Monitoramento contínuo:** métricas como acurácia, PSI, KS, AUC e estabilidade das variáveis.
- Como garantir que os clusters continuem representando o comportamento dos clientes?
 - Criamos um **painel de monitoramento** para acompanhar mudanças nos clusters ao longo dos meses.
- Implementamos métricas como **Silhouette Score** e **PSI (Population Stability Index)** para medir a estabilidade da segmentação.
- Detectamos que, em alguns períodos promocionais, clientes do **Cluster 3** começaram a usar mais cupons, o que indicou a necessidade de reavaliar os clusters.

 A segmentação não é um processo estático. O comportamento do cliente muda e a modelagem precisa acompanhar.

Soft Skills Fundamentais

Traduzir o Problema de Negócio para uma Solução em Dados


Desafio: o time de marketing queria entender melhor o perfil dos clientes que usam cupons para otimizar campanhas promocionais. Mas... como transformar isso em um problema estatístico?

Solução proposta:

- **Definição clara do problema:** identificar padrões de comportamento entre clientes que usam e não usam cupons.
- **Transformação em um problema de dados:** escolhemos **clustering** como abordagem, pois o objetivo era segmentação.
- **Seleção das variáveis:** Criamos métricas relevantes como frequência de compras, ticket médio e % de compras com cupons.

Traduzir o Problema de Negócio para uma Solução em Dados

Dica: Um estatístico que sabe estruturar um problema de negócio para análise de dados gera muito mais valor para a empresa!

 A habilidade de **traduzir um desafio de negócio em um problema estatístico bem definido** é essencial para qualquer profissional de dados

Traduzir os Resultados Técnicos para o Negócio

Desafio: após rodar o modelo de clustering, tínhamos três grupos distintos. Como comunicar isso para o time de marketing de forma útil e clara?

Solução proposta:

- **Evitar jargões estatísticos:** Em vez de falar de centroides e métricas de distância, explicamos os clusters em termos de comportamento do cliente.
- **Foco na ação:** Em vez de apenas mostrar gráficos, sugerimos estratégias para cada perfil de cliente (ex: limitar cupons para os "Caçadores de Promoção").
- **Visualização eficaz:** Criamos dashboards simples com insights acionáveis e não apenas métricas estatísticas.

Traduzir os Resultados Técnicos para o Negócio

Dica: Saber comunicar insights de forma clara é o diferencial entre um analista técnico e um cientista de dados estratégico.




Resultados estatísticos só geram impacto quando são **compreendidos e aplicados** pelo negócio!

Pensamento Crítico

Validando se a Segmentação Faz Sentido

Estar próximo da área de negócios durante o desenvolvimento da solução (modelo) para realizar ajustes se necessário.


 Sempre validar os resultados com métricas e testes adicionais antes de apresentar conclusões.

Colaboração Multidisciplinar

Interação com Marketing e Risco

Desafio: Como alinhar objetivos entre diferentes áreas da empresa?

- O time de **marketing** queria maximizar a adesão aos cupons.
- O time de **risco** se preocupava com clientes que usavam cupons apenas para maximizar benefícios e tinham alto risco de inadimplência.
- **Solução:** Criamos duas segmentações paralelas:
 - Uma focada em **engajamento do cliente** (usando padrões de compras).
 - Outra baseada no **risco de inadimplência**, cruzando os clusters com scores de crédito.
 - Assim, cada time usou a análise dentro da sua própria estratégia sem conflitos.

 O estatístico precisa ser um **tradutor de dados**, facilitando a comunicação entre diferentes áreas.

Diferenciais

O que torna um estatístico diferenciado no mercado de ciência de dados

- **Base sólida em Estatística:** Modelos e análises bem fundamentadas garantem credibilidade e impacto real nos negócios.
- **Formação acadêmica como diferencial:** O Mestrado e Doutorado aprofundam a capacidade analítica e o pensamento crítico, mesmo sendo mais teóricos. Esses fundamentos fazem a diferença no mercado.
- **Capacidade de conectar teoria e prática:** Empresas valorizam profissionais que sabem aplicar conceitos estatísticos em problemas reais e tomar decisões baseadas em dados.

O que eu gostaria de ter sabido antes de entrar no mercado

- **Pressão para Entregar Resultados Rápidos:** O mercado financeiro exige soluções rápidas, o que pode ser um desafio para quem está acostumado a um ritmo acadêmico com um rigor científico maior. “O feito é melhor que o bem feito”.
- **Complexidade de Dados:** A variedade e volume de dados no mercado são enormes e exigem uma mentalidade ágil para processar e aplicar os modelos.
- **Comunicação Técnica:** Traduzir modelos técnicos para algo que stakeholders não técnicos consigam entender é uma habilidade que aprendi a valorizar.

Dicas Finais para Carreira em Ciência de Dados

- **Técnicas:** Mantenha-se atualizado com as ferramentas de análise de dados, como SQL, PySpark, e os fundamentos de machine learning. Mas também procure entender a base teórica por trás de cada técnica.
- **Interpessoais:** Desenvolva *soft skills*, como negociação, comunicação e adaptabilidade. A habilidade de trabalhar em equipe e de se adaptar rapidamente às mudanças do mercado é essencial para o sucesso.

Obrigado!

magnotairone@gmail.com

<https://www.linkedin.com/in/magnotairone/>