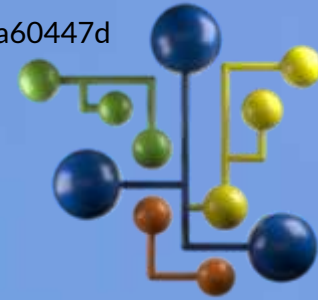




Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



Big Data Analytics com R e Microsoft Azure Machine Learning



Big Data Analytics com R e Microsoft Azure Machine Learning

Análise de Regressão com Linguagem R e Azure Machine Learning

Seja Bem-Vindo(a)!



Análise de Regressão com Linguagem R e Azure Machine Learning

Hora de consolidar tudo que estudamos e colocar o conhecimento em prática!





Análise de Regressão com Linguagem R e Azure Machine Learning

Processo de Data Science Para Análise de Big Data (Big Data Analytics)

- Compreender o Problema a ser Resolvido
- Coletar os Dados
- Explorar, Limpar e Preparar os Dados
- Selecionar e Transformar as Variáveis
- Construir, Testar, Avaliar e Otimizar o Modelo
- Contar a História dos Dados



Big Data Analytics com R e Microsoft Azure Machine Learning

O Que é Feature Selection?

Seja Bem-Vindo(a)!



O Que é Feature Selection?

Feature Selection





O Que é Feature Selection?

Por Que Usamos Feature Selection?



- Simplificação do modelo, para facilitar sua interpretação.
- Redução do tempo de treinamento do modelo.
- Melhora da generalização do modelo, evitando overfitting.



O Que é Feature Selection?

Feature Selection





O Que é Feature Selection?

Que variáveis (features) presentes em
nosso conjunto de dados, devem ser
usadas na criação do modelo?



O Que é Feature Selection?

As técnicas de feature selection automatizam a seleção das variáveis com maior potencial para variáveis preditoras.



O Que é Feature Selection?

Feature Selection é uma espécie de filtro, que remove do seu dataset as variáveis que não serão úteis para a criação do modelo preditivo.



O Que é Feature Selection?

O principal objetivo ao usar técnicas de Feature Selection é criar um modelo preditivo com a maior precisão possível e que seja generalizável.



O Que é Feature Selection?

Existem diversos métodos para Feature Selection

Teste do Qui-quadrado

Coeficientes de Correlação

Algoritmos de Eliminação Recursiva

Algoritmos de Regularização (LASSO, Elastic Net, Ridge Regression)



O Que é Feature Selection?



Atenção



O Que é Feature Selection?

Feature Selection



Redução de Dimensionalidade



O Que é Feature Selection?

Redução de Dimensionalidade

Principal Component Analysis (PCA)

Singular Value Decomposition (SVD)



O Que é Feature Selection?

Antes de aplicar Feature Selection,
diversas perguntas devem ser respondidas.



O Que é Feature Selection?

Antes de aplicar Feature Selection, diversas perguntas devem ser respondidas:

- Suas variáveis são mensuráveis?
- Você encontrou interdependência entre as variáveis?
- Você tem conhecimento sobre a área de negócio que gerou os dados?
- Sabe identificar as variáveis mais relevantes dentro do seu conjunto de dados?
- A análise exploratória dos dados encontrou "sujeira" nos seus dados?



O Que é Feature Selection?

As técnicas de Feature Selection basicamente calculam o nível de significância de cada variável e eliminam aquelas com significância mais baixa.



O Que é Feature Selection?

Veremos duas formas de fazer isso

Usando o módulo **Filter Based Feature Selection** do Azure ML

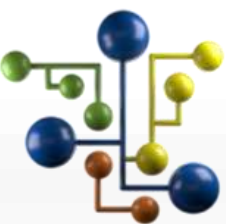
Criando um **modelo randomForest** para calcular a significância de cada variável, usando R



Big Data Analytics com R e Microsoft Azure Machine Learning

Storytelling

Seja Bem-Vindo(a)!



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Storytelling

Storytelling



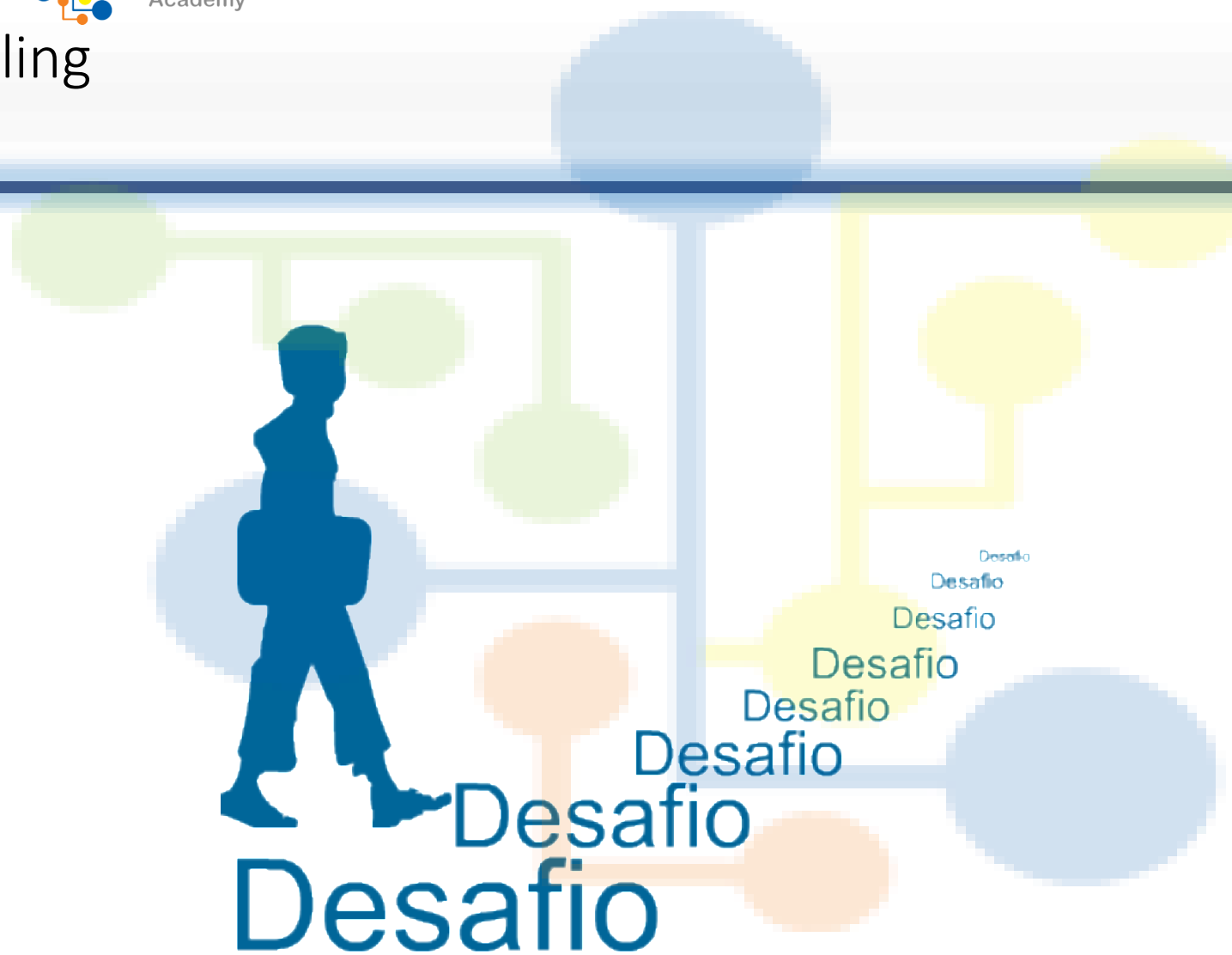


Storytelling

Como resumir todo um trabalho de análise e explicar isso a quem não conhece Data Science, Estatística, Programação ou Machine Learning?



Storytelling





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Forrest Gump





Storytelling

Não utilize linguagem técnica em apresentações executivas.



Storytelling

Use apenas um ou dois gráficos para explicar seus resultados (use tabelas se necessário).



Storytelling

Tente construir uma linha do tempo, explicando como os dados geraram o resultado preditivo esperado.



Storytelling

Não use termos como "variáveis" e sim
"atributos" ou "características".



Storytelling

Documente todo seu trabalho, pois isso servirá de apoio para que você possa contar a história dos dados.



Storytelling

Os tomadores de decisão não estão interessados em saber se você usou um algoritmo de regressão ou de classificação e nem como você otimizou seu modelo. Eles querem saber o que precisam fazer para aumentar as vendas e o faturamento!



Storytelling

Responda o que é importante para eles e não o que é importante para você!



Storytelling

Com o objetivo de negócio que foi traçado no início, mostre o que deve ser feito, com base na análise realizada. O processo de análise não interessa aos tomadores de decisão, mas sim o resultado final!



Storytelling

A fase de Análise Exploratória dos Dados pode apresentar insights preciosos. Use isso ao seu favor e guarde todos os resultados intermediários do seu processo de análise!



Storytelling

Storytelling é uma arte e requer prática para ser exercida com maestria!

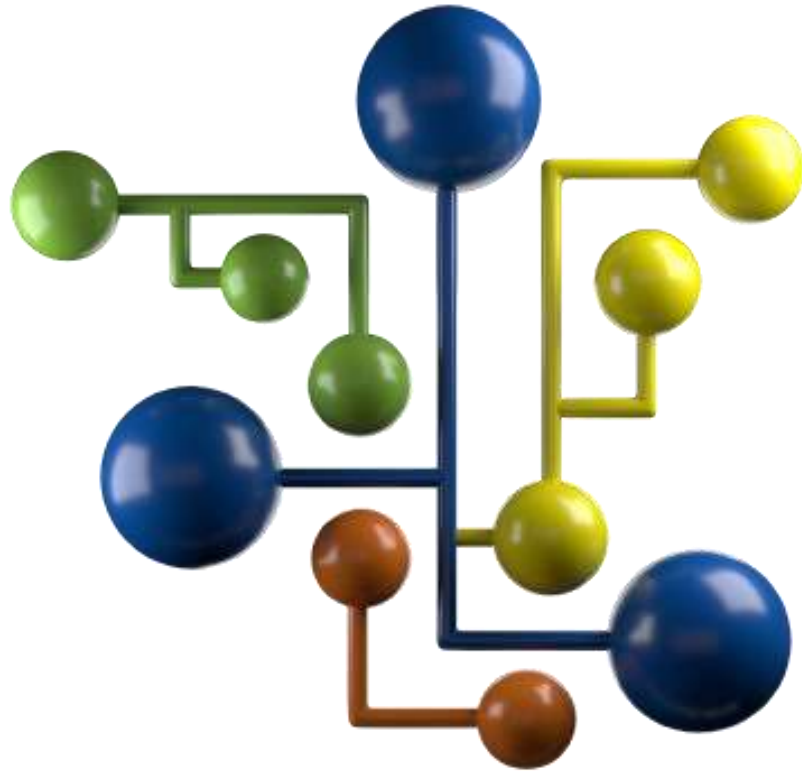


Storytelling

No curso de Visualização de Dados e Design de Dashboards, você aprenderá a realizar uma apresentação executiva a partir do resultado da sua análise!



Muito Obrigado por Participar!



Tenha uma Excelente Jornada de Aprendizagem.

Equipe Data Science Academy

