

1500

FIFTEEN HUNDRED

transformando
pessoas e
organizações



1500

FIFTEEN HUNDRED

Curso 1500fh Formação em Dados
e Analytics

Modulo 1 – Aula 2

- Visão inicial do que é um algoritmo (para que serve, qual uso, como aplicar etc)

1500
FIFTEEN HUNDRED

Caso 2 - Questões

Continuamos com nosso desvio na Margem de Contribuição...

Uma alternativa levantada pelo corpo diretivo é aumentar as vendas de banda larga neste momento de covid. Porém, nossa equipe de call center é limitada, assim como a verba de marketing. Precisamos focar nossos esforços e aumentar nossa eficácia de venda.

Temos esta base de dados com informações dos clientes.

Questões

- Quem são os clientes de maior propensão?

Qual pergunta precisamos responder?

- Dadas as características de um cliente, ele vai comprar banda larga, sim ou não?

Como?

O QUE É UM ALGORITMO?

- Sequência de instruções finitas com um objetivo

-> $X = 3;$

-> $Y = 5;$

-> $C = X + Y;$

C é 8

O QUE É APRENDIZADO DE MÁQUINA?

O QUE É APRENDIZADO DE MÁQUINA?

- Área de pesquisa da Inteligência Artificial que visa o desenvolvimento de algoritmos capazes de aprender a executar uma tarefa

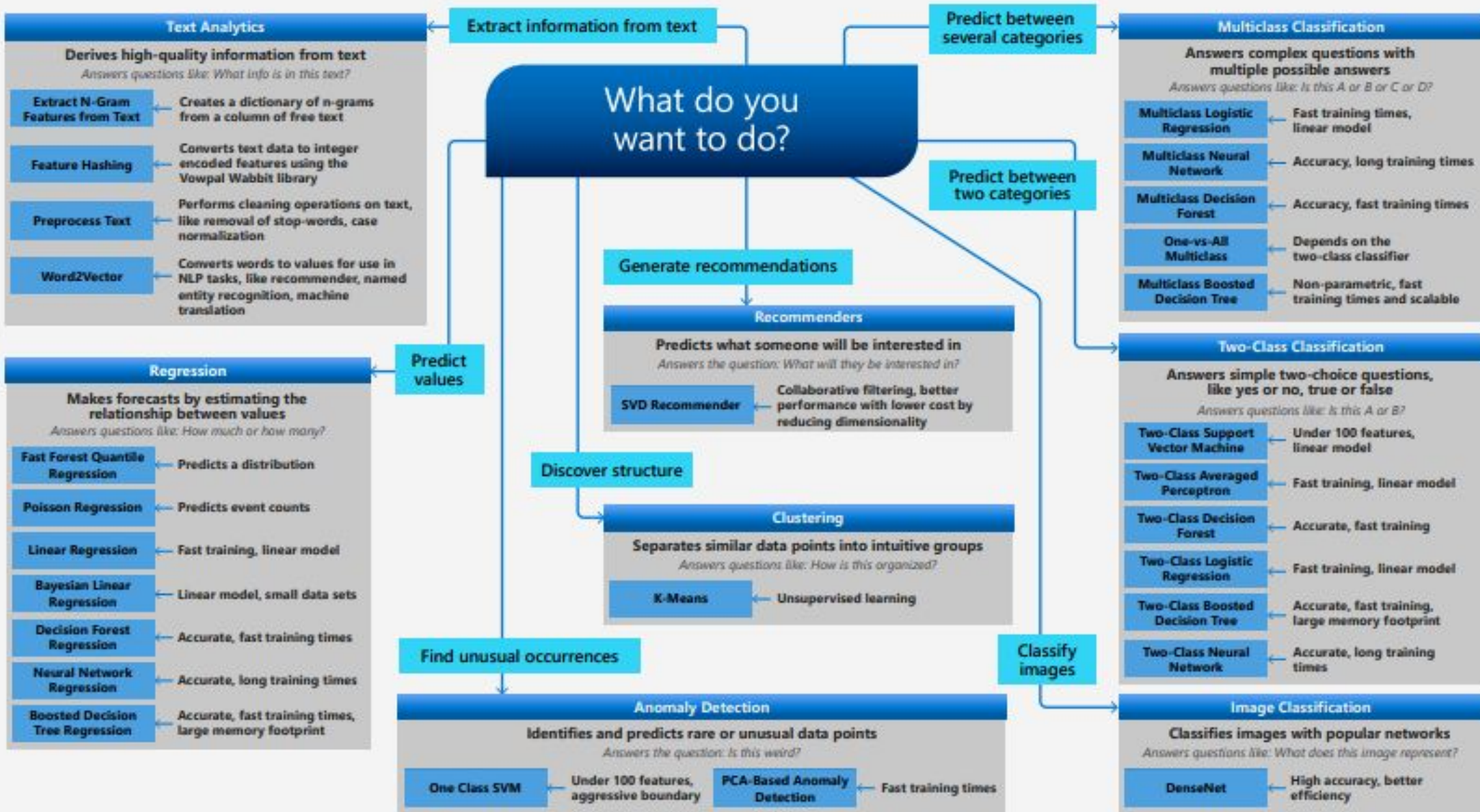
"Da aos computadores a capacidade de aprender sem serem explicitamente programados." - **Arthur Samuel**



Microsoft Azure Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

1500

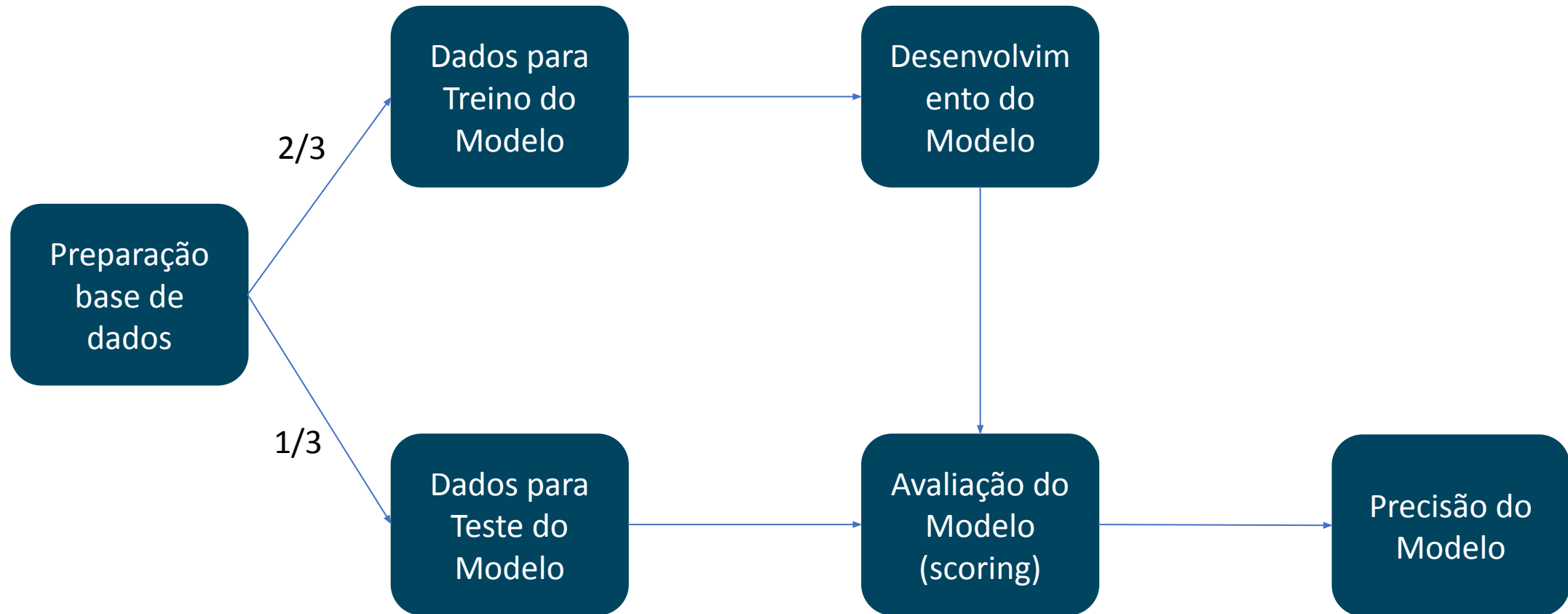


- O objetivo de um **modelo de classificação** é sempre explicar ou prever a ocorrência de determinado evento em função de um conjunto de variáveis, que podem ser categóricas ou não.
- É importante observar que a variável dependente é de natureza binária e exige resultados que possam ser interpretados em termos de probabilidade.
- Quais exemplos de perguntas que este tipo de modelo responde?

Exemplos de uso de algoritmos de classificação:

- Aluno ser aprovado ou reprovado num exame
- Um paciente vir a óbito ou sobreviver a determinada condição
- Um candidato a um posto de trabalho ser contratado ou não
- Um produto ser aceito ou barrado pelo controle de qualidade
- Um cliente cancelar ou não um pedido
- Um gerente obter êxito ou fracassar numa negociação
- Um fornecedor aceitar ou rejeitar uma proposta
- Um cliente tornar-se inadimplente ou não
- Uma cidade sofrer um ataque terrorista ou não
- Uma cidade passar por problemas de abastecimento ou não
- Uma empresa enfrentar greves ou não

PROCESSO CONSTRUÇÃO E SELEÇÃO DO MODELO



Preparação de dados para classificação

- Seleção de variáveis
 - Remoção dos atributos irrelevantes ou redundantes
- Preparação de dados
 - Pré-processamento dos dados para reduzir o ruído e tratar os valores ausentes
- Transformação dos dados
 - Generalização e/ou normalização dos dados
- Balanceamento
 - Equalização do dataset de treino

Preparação de dados para classificação


- Seleção de variáveis
 - Remoção dos atributos irrelevantes ou redundantes
- Preparação de dados
 - Pré-processamento dos dados para reduzir o ruído e tratar os valores ausentes
- Transformação dos dados
 - Generalização e/ou normalização dos dados
- Balanceamento
 - Equalização do dataset de treino

Treino e seleção do Modelo

- Via Azure Machine Learning Studio (Ferramenta)
- Testar pelo menos 2 algoritmos diferentes
- Escolher o que apresentar melhor resultado

Métricas

Métricas para avaliação e escolha do melhor modelo

		PREVISÃO		
		0	1	
RESULTADO REAL	0			CONFUSION MATRIX
	1			

Métricas

Métricas para avaliação e escolha do melhor modelo

		PREVISÃO	
		0	1
RESULTADO REAL	0	VERDADEIRO NEGATIVO	FALSO POSITIVO
	1	FALSO NEGATIVO	VERDADEIRO POSITIVO

CONFUSION MATRIX

- **Verdadeiro Positivo:** quando o modelo prevê corretamente um resultado positivo
- **Verdadeiro Negativo:** quando o modelo prevê corretamente um resultado negativo
- **Falso Positivo:** quando o modelo prevê erradamente um resultado positivo
- **Falso Negativo:** quando o modelo prevê erradamente um resultado negativo

Métricas

Métricas para avaliação e escolha do melhor modelo

		PREVISÃO	
		0	1
RESULTADO REAL	0	VERDADEIRO NEGATIVO	FALSO POSITIVO
	1	FALSO NEGATIVO	VERDADEIRO POSITIVO

$$\text{Accuracy} = \frac{vp + vn}{vp + vn + fp + fn}$$

$$\text{Precision Score} = \frac{vp}{vp + fp}$$

$$\text{Recall Score} = \frac{vp}{vp + fn}$$

$$\text{F1 score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

Métricas

EXEMPLO

		PREVISÃO	
		0	1
RESULTADO REAL	0	100	10
	1	20	50

$$\text{Accuracy} = (vp + vn) / (vp + vn + fp + fn)$$

$$\text{Precision Score} = vp / (vp + fp)$$

$$\text{Recall Score} = vp / (vp + fn)$$

$$\text{F1 score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

Métricas

EXEMPLO

		PREVISÃO	
		0	1
RESULTADO REAL	0	100	10
	1	20	50

$$\text{Accuracy} = \frac{vp + vn}{vp + vn + fp + fn} = \frac{(50 + 100)}{(50 + 100 + 10 + 20)} = 0,8333$$

$$\text{Precision Score} = \frac{vp}{vp + fp} = \frac{50}{(50 + 10)} = 0,83333$$

$$\text{Recall Score} = \frac{tp}{tp + fn} = \frac{50}{(50 + 20)} = 0,714$$

$$\text{F1 score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} = \frac{2 * 0,8333 * 0,714}{(0,8333 + 0,714)} = 0,769$$

Métricas

Accuracy

- Razão de amostras corretas: $vp + vn / vp + vn + fp + fn$, onde:

vp – verdadeiros positivos
vn – verdadeiros negativos
fp – falsos positivos
fn – falsos negativos

Precision Score

- Razão $vp / (vp + fp)$, onde vp é o número de verdadeiros positivos e fp o número de falsos positivos.
- Capacidade do modelo não rotular como positiva uma amostra negativa

Recall Score

- Razão $vp / (vp + fn)$, onde vp é o número de verdadeiros positivos e fn o número de falsos negativos.
- Capacidade do modelo encontrar todas as amostras positivas

F1 Score

- Média ponderada de Precision e Recall. F1 atinge o melhor valor em 1 e o pior valor em 0.
- $F1 = 2 * (precision * recall) / (precision + recall)$

- **Para o caso 2, qual métrica considerar para a escolha do melhor modelo?**

- **AZURE MACHINE LEARNING STUDIO**

- Apresentação da solução:
 - Declarar necessidade do cliente
 - Definir uma forma para atender esta necessidade
 - Descrever o algoritmo, seu uso, estrutura de dados, qualidade e resultados (testar pelo menos dois modelos diferentes e justificar a escolha através dos resultados obtidos)



1500

FIFTEEN HUNDRED

Obrigado!

Paulo Seixas

pseixas@1500fh.com