

Curso Aprendizado de Máquina

O processo de Classificação

- Os métodos gerais que os modelos de tipo sklearn implementam são:
- Analisar os dados
- Ajustar um modelo
- Prever classes
- Obtenha precisão

Os dados

- Os dados em geral podem vir de um DataFrame Pandas, que pode ser extraído de arquivos do excel, csv, sql entre outros.
- No aprendizado supervisionado em geral os dados são divididos em:
- Atributos de predição:
 - São os dados de entrada do modelo
 - Em geral são dados com duas colunas ou mais
- Atributo alvo:
 - São os dados que desejamos prever e que o algoritmo irá usar para o aprendizado
 - Na classificação binária eles possuem apenas valores binários, como 0 e 1 ou aprovado e reprovado

Um exemplo

- Vamos imaginar um restaurante que possui os seguintes dados:
 - R_{ID} : identificação de uma refeição
 - TEMPO: tempo médio de preparo da refeição no último mês
 - PREÇO: preço das refeições servidas no último mês
 - SAT: satisfeito
 - INS: insatisfeito
- O dono do restaurante faz a seguinte pergunta:
 - Como é possível prever a satisfação dos clientes com base no tempo de preparo de uma refeição e seu preço ?
 - Para isso ele coletou os dados TEMPO, PREÇO e SATISFAÇÃO e deseja que você elabore uma análise para ele

Os dados coletados

RID	TEMPO	PREÇO	CLASSE
R1	28	R\$ 60.00	SAT
R2	16	R\$ 75.00	SAT
R3	24.4	R\$ 60.00	SAT
R4	36	R\$ 105.00	SAT
R5	32	R\$ 45.00	SAT
R6	24	R\$ 150.00	SAT
R7	8	R\$ 90.00	INS
R8	16	R\$ 45.00	INS
R9	24.4	R\$ 30.00	INS
R10	24	R\$ 64.50	INS
R11	20	R\$ 15.00	INS
R12	40	R\$ 30.00	INS

Transformação dos dados

- Os atributos descritivos em um conjunto de dados assumem valores que contribuem para a caracterização de um exemplar.
- Porém, essa caracterização pode assumir valores de diferentes naturezas.
- É comum, por exemplo, o uso de diferentes tipos de dados para aquisição de dados armazenadas em grandezas diferentes.
- Por exemplo um site pode apresentar os valores de um produto em reais e outro em dólares, ou os dados de temperatura em uma fonte estão em graus celsius (°C) e em Farenheith (°F).
- Ou ainda você pode ter dados de data no formato brasileiro DD-MM-YYYY e no formato americano MM-DD-YYYY.
- Ainda, uma base de dados transacional que armazena características de objetos ou transações e possui atributos com valores pertinentes a diferentes domínios.

Normalização e Conversão

- Para amenizar os efeitos causados por situações como essas, é necessário aplicar procedimentos de **transformação de dados** (também conhecidos como *data transformation*).
- Esses procedimentos abrangem a **normalização de dados** e a **conversão de dados**.

Normalização

- A técnica de normalização mais comum é a *normalização min-max*, que altera os valores extremos do conjunto de valores sob normalização e organiza os demais valores dentro do novo intervalo de domínio;.
- *Outra técnica de normalização é a z-score*, por meio da qual os valores são reorganizados pela média e desvio-padrão do conjunto de valores original.

Normalização Min Max

- A *normalização min-max* é uma das formas mais simples de normalização aplicada em problemas de mineração de dados.
- Quando aplicada a cada atributo de um conjunto de dados, permite que todos eles sejam organizados em uma mesma escala de valores.
- Para normalizar os valores com um mínimo de 0 e um máximo de 1 esse procedimento é formalizado como:

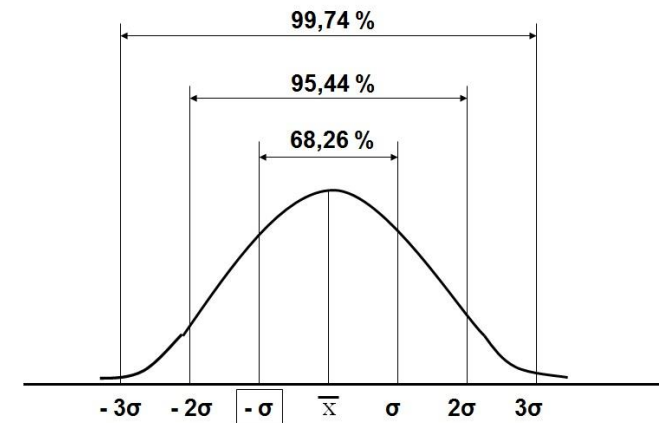
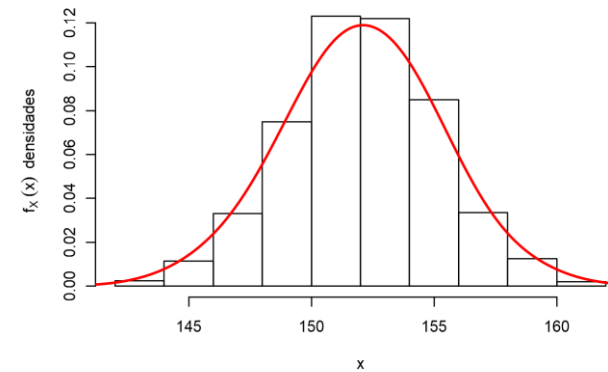
$$Valor_{normalizado} = \frac{Valor_{não\ normalizado} - Valor_{mínimo}}{Valor_{máximo} - Valor_{mínimo}}$$

- Por exemplo para TEMPO temos
- Valor Máximo 40 min
- Valor Mínimo 8 min
- Desejamos normalizar o tempo de uma refeição que demorou 16 min
- $T_n = \frac{16 - 8}{40 - 8} = \frac{8}{32} = \frac{1}{4} = 0.25$
- Para PREÇO temos
- Valor Máximo R\$ 150,00
- Valor Mínimo R\$ 15,00
- Desejamos normalizar o preço de uma refeição que custou R\$ 60,00
- $P_n = \frac{60.00 - 15.00}{150.00 - 15.00} = \frac{45.00}{135.00} = \frac{9}{27} = 0.333$

Normalização z-score

- A *normalização z-score* também permite que os valores sejam padronizados.
- No entanto, isso ocorre com base na distribuição dos valores de cada atributo.
- Com esta normalização, cada valor do atributo é determinado pela posição que ocupa em relação aos demais, em uma distribuição simétrica.

$$Valor_{normalizado} = \frac{Valor_{não\ normalizado} - Média}{Desvio\ Padrão}$$



Os dados de tempo normalizado (min max)

RID	TEMPO NORMALIZADO	TEMPO
R1	0.625	28
R2	0.250	16
R3	0.513	24.4
R4	0.875	36
R5	0.750	32
R6	0.500	24
R7	0.000	8
R8	0.250	16
R9	0.513	24.4
R10	0.500	24
R11	0.375	20
R12	1.000	40

Os dados de PREÇO normalizado (min max)

RID	PREÇO NORMALIZADO	PREÇO
R1	0.333	R\$ 60.00
R2	0.444	R\$ 75.00
R3	0.333	R\$ 60.00
R4	0.667	R\$ 105.00
R5	0.222	R\$ 45.00
R6	1.000	R\$ 150.00
R7	0.556	R\$ 90.00
R8	0.222	R\$ 45.00
R9	0.111	R\$ 30.00
R10	0.367	R\$ 64.50
R11	0.000	R\$ 15.00
R12	0.111	R\$ 30.00

Conversão dos dados

- Uma conversão de dados pode ser necessária, pelo menos, sob duas formas: conversão de valores numéricos para categóricos e conversão de valores categóricos para numéricos.
- O primeiro caso, também conhecido como discretização , pode ser implementado por meio da criação de categorias.
- Já o segundo caso é mais bem nomeado como codificação.
- Trata-se de representar os dados categóricos usando números, sem, no entanto, permitir que os algoritmos que analisarão tais valores possam estabelecer relações entre eles quando elas não existem originalmente (por exemplo, relações de ordem, prioridade, importância ou quantidade).

Conversão de dados

Categórico

- Sexo
 - [masculino, feminino]
- Resultado
 - [positivo, negativo]
- Aprovação
 - [reprovado, recuperação, aprovado]
- Animal 1:
 - [gato, cachorro]
- Animal 2:
 - [gato, cachorro, peixe]

Numérico

- Sexo
 - [0, 1]
- Resultado
 - [0, 1]
- Aprovação
 - [0, 1, 2]
- Animal 1:
 - [0, 1]
- Animal 2:
 - [0, 1, 2]

Os dados de CLASSE convertidos

RID	CLASSE	CLASSE CONVERTIDA
R1	SAT	1
R2	SAT	1
R3	SAT	1
R4	SAT	1
R5	SAT	1
R6	SAT	1
R7	INS	0
R8	INS	0
R9	INS	0
R10	INS	0
R11	INS	0
R12	INS	0

Os dados coletados, normalizados e convertidos

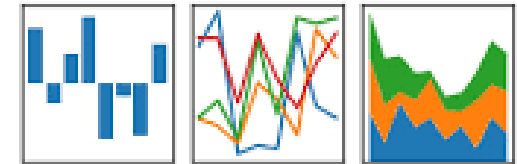
RID	TEMPO NORMALIZADO	PREÇO NORMALIZADO	CLASSE CONVERTIDA
R1	0.625	0.333	1
R2	0.250	0.444	1
R3	0.513	0.333	1
R4	0.875	0.667	1
R5	0.750	0.222	1
R6	0.500	1.000	1
R7	0.000	0.556	0
R8	0.250	0.222	0
R9	0.513	0.111	0
R10	0.500	0.367	0
R11	0.375	0.000	0
R12	1.000	0.111	0

Análise dos Dados com pandas

- É conveniente que analisar dados antes de iniciar uma análise com aprendizado de máquina
- Para isso vamos usar a biblioteca pandas
- Para importar a biblioteca pandas usamos a seguinte linha de código:
- **import pandas as pd**
- Onde pd é uma abreviação que podemos usar toda vez que quisermos chamar uma função do pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Pandas



Importando dados do excel

- O pandas permite a importação de dados diretamente do excel usando a seguinte linha de código:
- `df = pd.read_excel('Dados_Satis.xlsx')`
- `df` é uma variável onde vamos armazenar o DataFrame
- `Pd` é a abreviação, com ela o python sabe que estamos usando uma função do pandas
- `'Dados_Satis.xlsx'` é uma string com o endereço do arquivo a converter e pode ser uma url.
- Os dados aparecem como ao lado

1	df						
	RID	TEMPO_N	PREÇO_N	CLASSE	TEMPO	PREÇO	CLASSE_C
0	R1	0.6250	0.333333	SAT	60.0	28.0	1
1	R2	0.2500	0.444444	SAT	75.0	16.0	1
2	R3	0.5125	0.333333	SAT	60.0	24.4	1
3	R4	0.8750	0.666667	SAT	105.0	36.0	1
4	R5	0.7500	0.222222	SAT	45.0	32.0	1
5	R6	0.5000	1.000000	SAT	150.0	24.0	1
6	R7	0.0000	0.555556	INS	90.0	8.0	0
7	R8	0.2500	0.222222	INS	45.0	16.0	0
8	R9	0.5125	0.111111	INS	30.0	24.4	0
9	R10	0.5000	0.366667	INS	64.5	24.0	0
10	R11	0.3750	0.000000	INS	15.0	20.0	0
11	R12	1.0000	0.111111	INS	30.0	40.0	0

Estatísticas

- Você pode ver uma estatística dos dados com o seguinte comando
- `df.mean()` – exibe a média das variáveis
 - TEMPO_N 0.51
 - PREÇO_N 0.36
 - TEMPO 24.40
 - PREÇO 64.12
- Sabemos então que o preço médio de uma refeição é R\$ 64.12
- E o tempo médio é 24.4 min

Estatísticas

- Você pode ver uma estatística dos dados com o seguinte comando
- `df.describe()` – exibe um sumário estatístico das variáveis
- Sabemos então que o preço médio de uma refeição é R\$ 64.12
- E o tempo médio é 24.4 min

Estatísticas do DataFrame

	TEMPO_N	PREÇO_N	PREÇO	TEMPO	CLASSE_C
count	12.0000	12.0000	12.0000	12.0000	12.0000
mean	0.51250	0.36389	64.1250	24.4000	0.50000
std	0.27925	0.27663	37.3449	8.93614	0.52223
min	0.00000	0.00000	15.0000	8.00000	0.00000
25%	0.34375	0.19444	41.2500	19.0000	0.00000
50%	0.50625	0.33333	60.0000	24.2000	0.50000
75%	0.65625	0.47222	78.7500	29.0000	1.00000
max	1.000000	1.00000	150.00	40.000	1.00000

- Onde:
- Count é a contagem dos dados
- Mean é o valor médio de cada atributo
- Std é o desvio padrão de cada atributo
- Min é o valor mínimo de cada atributo
- 25% é o valor dos primeiros 25 % dos valores de cada atributo (1º quartil)
- 50% é o valor dos primeiros 50 % dos valores de cada atributo (2º quartil)
- 75% é o valor dos primeiros 75 % dos valores de cada atributo (3º quartil)
- Max é o valor máximo de cada atributo

Entendendo as estatísticas medidas de posição

- Medidas de posição , ou medidas de tendência central , permitem encontrar os valores que orientam a análise dos dados no que diz respeito à sua localização, ou como a distribuição associada aos valores se comporta.
- As medidas de posição mais comuns no contexto de análise exploratória de dados são média aritmética, mediana e moda.
- Relacionadas com a mediana estão as medidas separatrizes percentis e quartis.

Entendendo as estatísticas - Média

- Para um conjunto de valores $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$, em que n é a quantidade de valores do conjunto, a **média** aritmética simples ($média(\mathbf{v})$ ou) é a soma dos valores do conjunto \mathbf{v} , dividida pela quantidade de valores presentes nesse conjunto. Formalmente, $média(\mathbf{v})$ é definida como:

$$média(v) = \frac{1}{n} \sum_{i=1}^n v_i$$

Estatísticas do DataFrame - Média

	TEMPO_N	PREÇO_N	PREÇO	TEMPO	CLASSE_C
count	12.0000	12.0000	12.0000	12.0000	12.0000
mean	0.51250	0.36389	64.1250	24.4000	0.50000
std	0.27925	0.27663	37.3449	8.93614	0.52223
min	0.00000	0.00000	15.0000	8.00000	0.00000
25%	0.34375	0.19444	41.2500	19.0000	0.00000
50%	0.50625	0.33333	60.0000	24.2000	0.50000
75%	0.65625	0.47222	78.7500	29.0000	1.00000
max	1.000000	1.00000	150.00	40.000	1.00000

- Onde:
- N é igual a 12
- A média de TEMPO_N é 0.5125
- A média de PREÇO_N é 0.3639
- A média de TEMPO é 64.12
- A média de PREÇO é 24.40

Entendendo as estatísticas – Mediana e Moda

- A **mediana** ($mediana(\mathbf{v})$) é o valor que divide a distribuição dos valores exatamente ao meio (podendo ser vista também como uma medida separatriz).
- Importante lembrar que tal valor não precisa estar presente no conjunto \mathbf{v} .
- A **moda**, por sua vez, é o valor mais frequente em um conjunto de valores. Ela é a única medida de posição que pode assumir mais de um valor.
- Essa situação ocorre quando dois ou mais valores aparecem no conjunto de valores \mathbf{v} com a mesma frequência, a máxima no conjunto.

Entendendo as estatísticas - Quartis

- Para um conjunto de valores $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$, em que n é a quantidade de valores do conjunto,.
- Os percentis e quartis são medidas separatrizes que dividem o conjunto de valores, ordenados de forma crescente, em partes tão iguais quanto possível.
- O percentil de ordem p determina os $p\%$ menores valores contidos em \mathbf{v} .
- **Em geral usamos 25%, 50% e 75%. Que são um quarto, metade treês quartos, definidos como 1º quartil, 2º quartil e 3º quartil**

Estatísticas do DataFrame

	TEMPO_N	PREÇO_N	PREÇO	TEMPO	CLASSE_C
count	12.0000	12.0000	12.0000	12.0000	12.0000
mean	0.51250	0.36389	64.1250	24.4000	0.50000
std	0.27925	0.27663	37.3449	8.93614	0.52223
min	0.00000	0.00000	15.0000	8.00000	0.00000
25%	0.34375	0.19444	41.2500	19.0000	0.00000
50%	0.50625	0.33333	60.0000	24.2000	0.50000
75%	0.65625	0.47222	78.7500	29.0000	1.00000
max	1.000000	1.00000	150.00	40.000	1.00000

- Onde:
- O valor do 1º quartil para TEMPO é 41.25
- O valor do 2º quartil para TEMPO é 60.00
- O valor do 3º quartil para TEMPO é 78.75
- O valor do 1º quartil para PREÇO é 19.00
- O valor do 2º quartil para TEMPO é 24.20
- O valor do 3º quartil para TEMPO é 29,00

Entendendo as estatísticas medidas de dispersão

- As medidas de posição são úteis para apresentar uma sumarização dos dados.
- No entanto, não são capazes de descrever a variação ou dispersão do conjunto de valores.
- Para esse fim, aplicam-se as medidas de dispersão , capazes de descrever o quanto os valores de um conjunto estão próximos ou distantes de uma medida central, como a média.
- As medidas mais comuns de dispersão ou variância dos dados são a amplitude, variância, desvio-padrão e coeficiente de variação.

Entendendo as estatísticas Amplitude

- Para um conjunto de valores \mathbf{v} , a medida da **amplitude** é a diferença entre o maior e o menor valor do conjunto, portanto: $\text{amplitude}(\mathbf{v}) = \max(\mathbf{v}) - \min(\mathbf{v})$.
- Embora a medida da amplitude seja simples, sua interpretação precisa ser feita com cuidado, pois ela pode ser influenciada por valores extremos, conhecidos como *outliers*

Entendendo as estatísticas desvio padrão

- A **variância** é uma medida de dispersão definida como a média dos quadrados das diferenças entre cada valor do conjunto **v** e a média desse conjunto.
- O **desvio-padrão** é a **raiz quadrada da variância** só faz sentido quando a média for usada como medida de posição (ou de centro).
- Ele é semelhante à medida de amplitude, com a diferença de que o cálculo do desvio-padrão usa todos os valores de um conjunto.
- O resultado dessa medida geralmente é usado para verificar a consistência de um fenômeno (um fenômeno é consistente quando o cálculo do desvio-padrão resulta em valores baixos).

$$var(v) = \frac{1}{n} \sum_{i=1}^n (v_i - média(v))^2 \quad desvpad(v) = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_i - média(v))^2}$$

Estatísticas do DataFrame

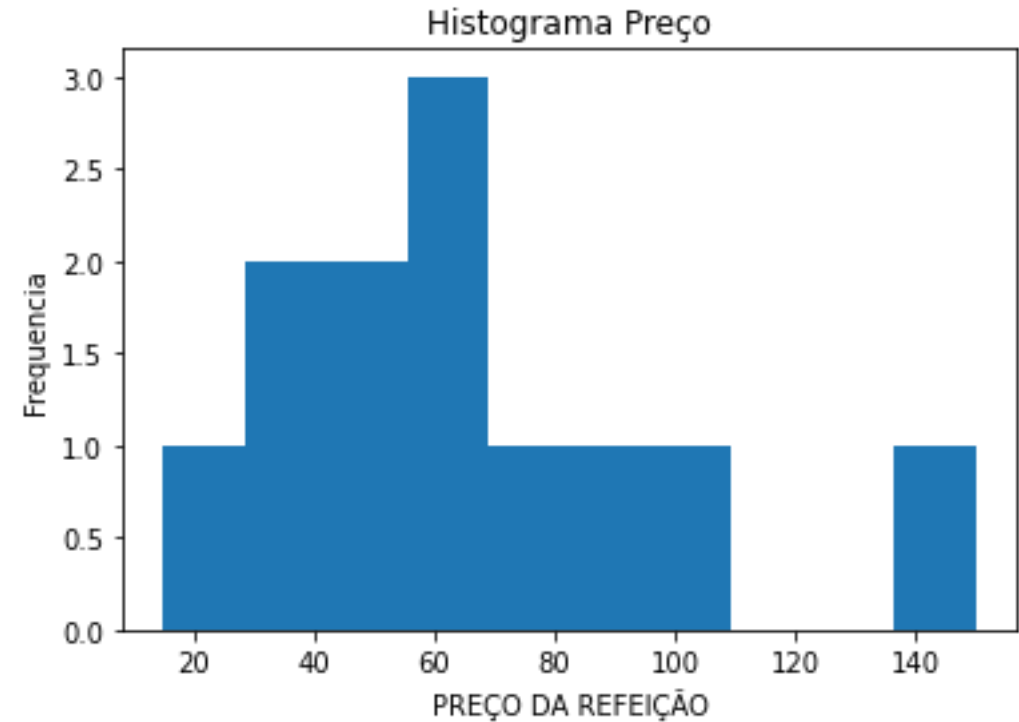
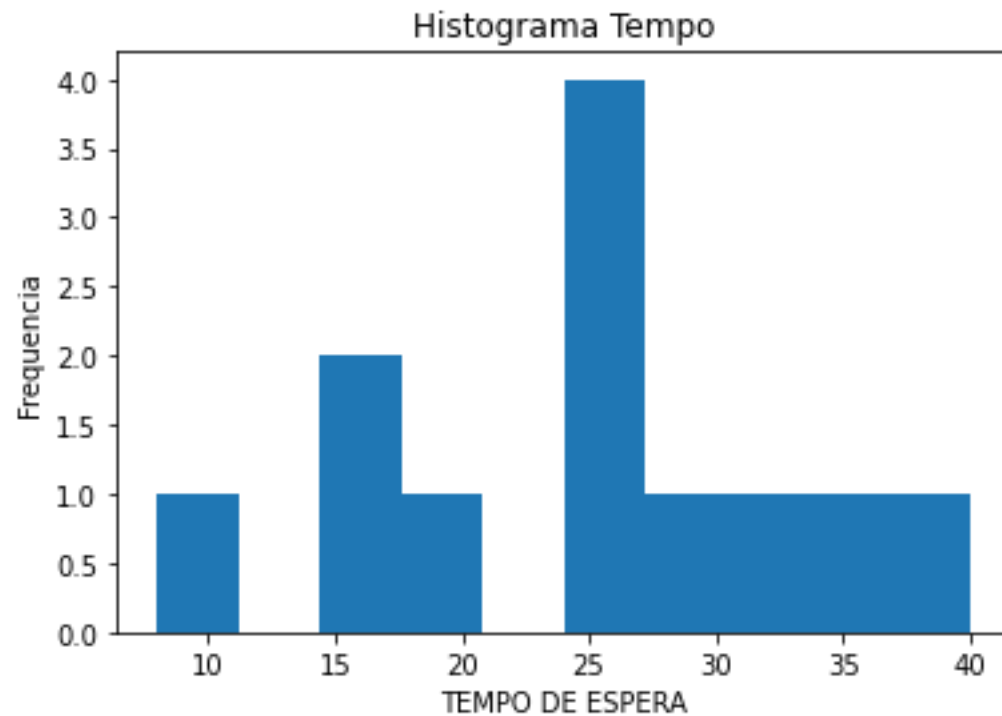
	TEMPO_N	PREÇO_N	PREÇO	TEMPO	CLASSE_C
count	12.0000	12.0000	12.0000	12.0000	12.0000
mean	0.51250	0.36389	64.1250	24.4000	0.50000
std	0.27925	0.27663	37.3449	8.93614	0.52223
min	0.00000	0.00000	15.0000	8.00000	0.00000
25%	0.34375	0.19444	41.2500	19.0000	0.00000
50%	0.50625	0.33333	60.0000	24.2000	0.50000
75%	0.65625	0.47222	78.7500	29.0000	1.00000
max	1.000000	1.00000	150.00	40.000	1.00000

- Onde:
- A amplitude de PREÇO é o máximo menos o mínimo ou seja 150-15 que é igual a 135
- A amplitude de TEMPO é o máximo menos o mínimo ou seja 40-8 que é igual a 32
- Ou seja os valores de TEMPO tem uma faixa de variação menor que PREÇO

Histogramas

- Um histograma , também conhecido como distribuição de frequências
- Pode ser apresentados em uma forma gráfica composta por colunas (ou barras).
- As colunas dizem respeito à divisão em faixas (geralmente uniforme) dos valores de uma variável de um conjunto de dados.

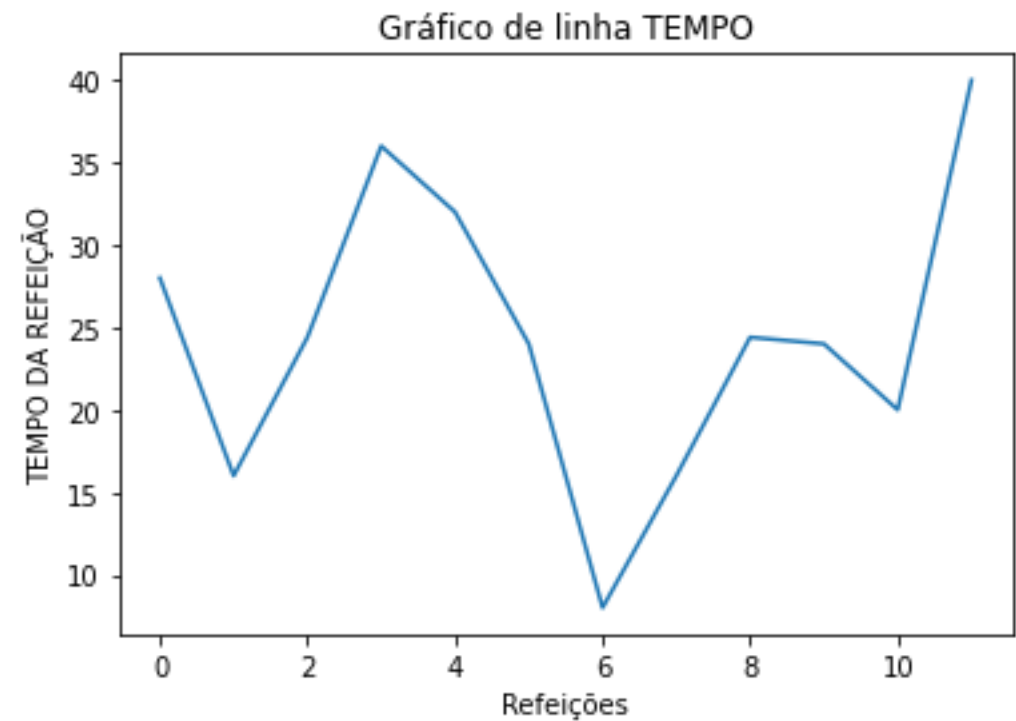
Histogramas de Preço e Tempo



Gráficos de Linhas

- Um gráfico de linhas pode ser usado para suportar a análise de atributos individualmente.
- Em geral, no eixo x de um gráfico de linhas, estão os índices dos exemplares do conjunto de dados.
- Também no eixo x de um gráfico de linhas podem estar os tempos em que foram coletados, caso o conjunto de dados venha de uma série temporal de valores.

Gráficos de linhas de Tempo e Preço



Graficos de Linha de TEMPO E PREÇO

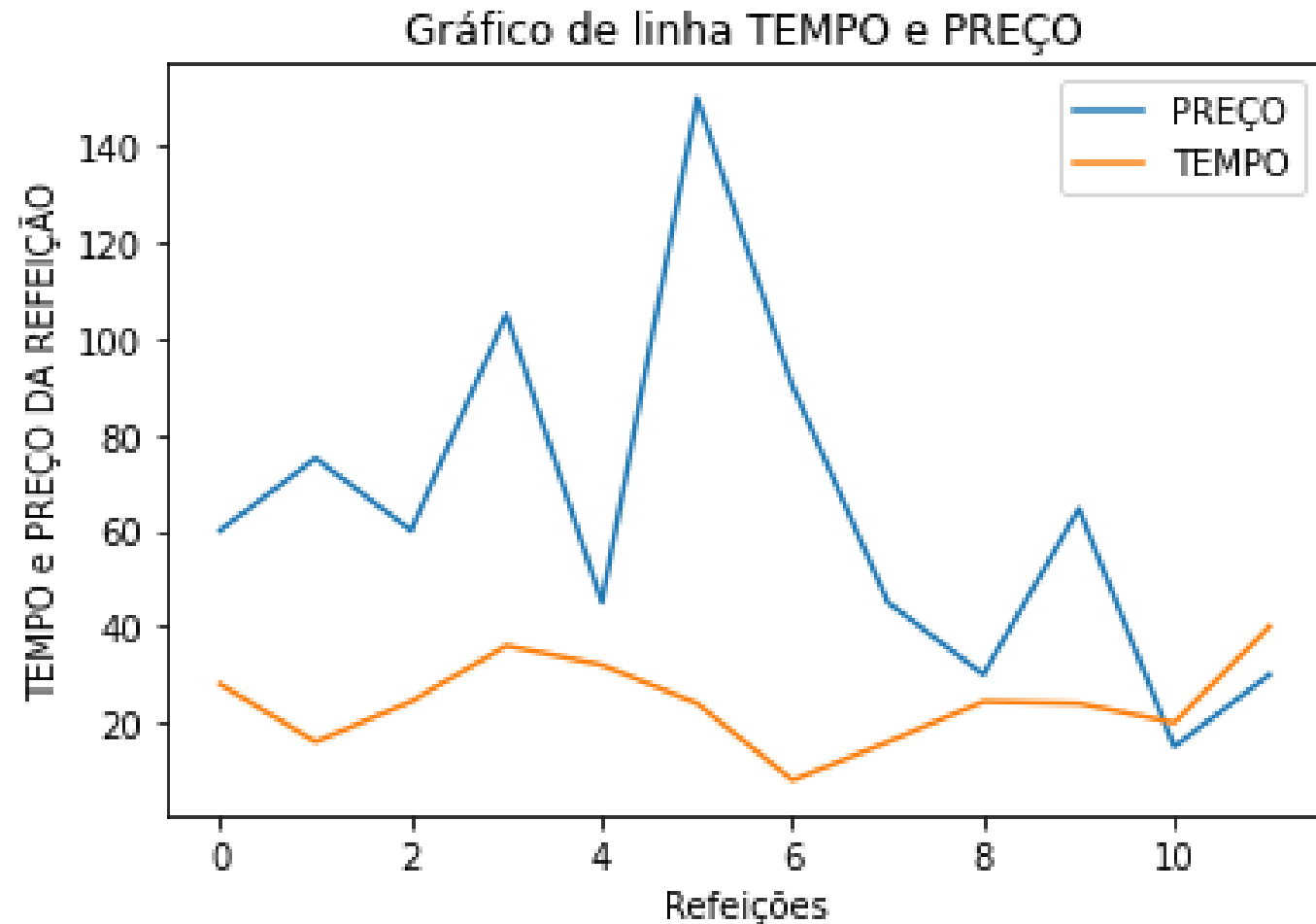


Gráfico de dispersão

- Um gráfico de dispersão (*scatter plot*), ou espalhamento, permite a visualização da distribuição dos exemplares de um conjunto de dados.
- Um gráfico de dispersão é útil se queremos saber como os dados se comportam em relação aos outros

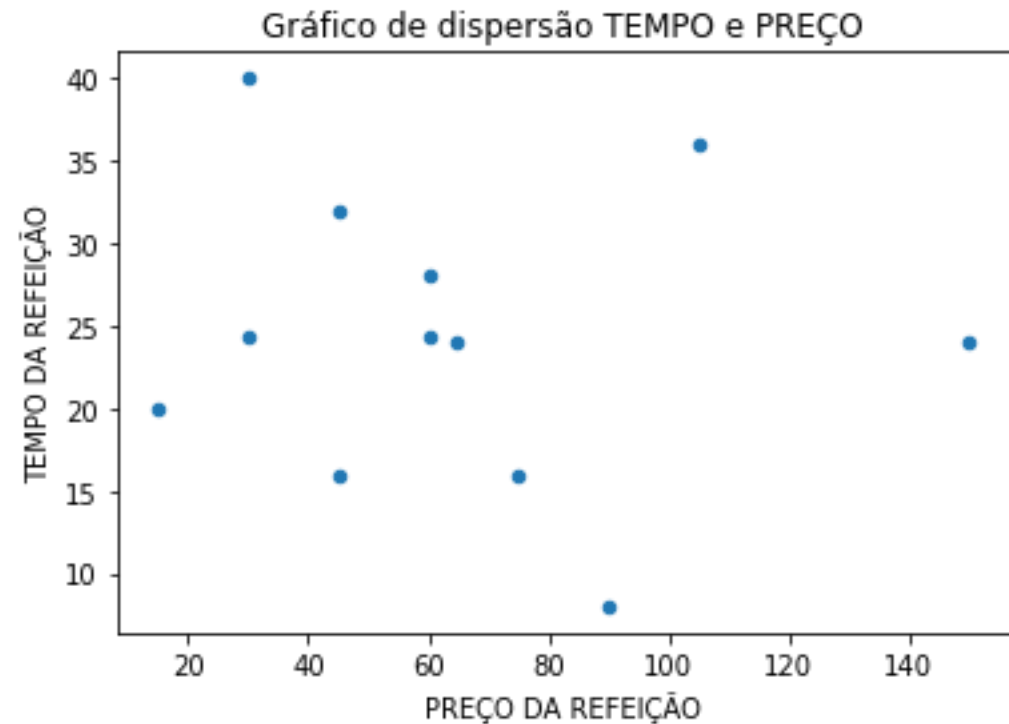
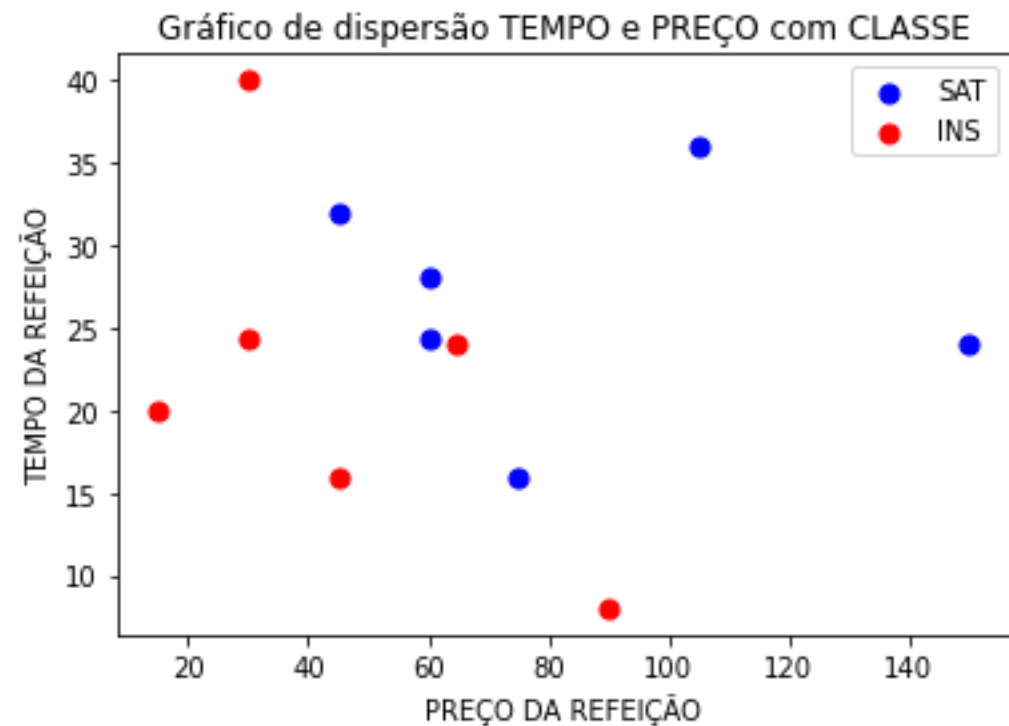
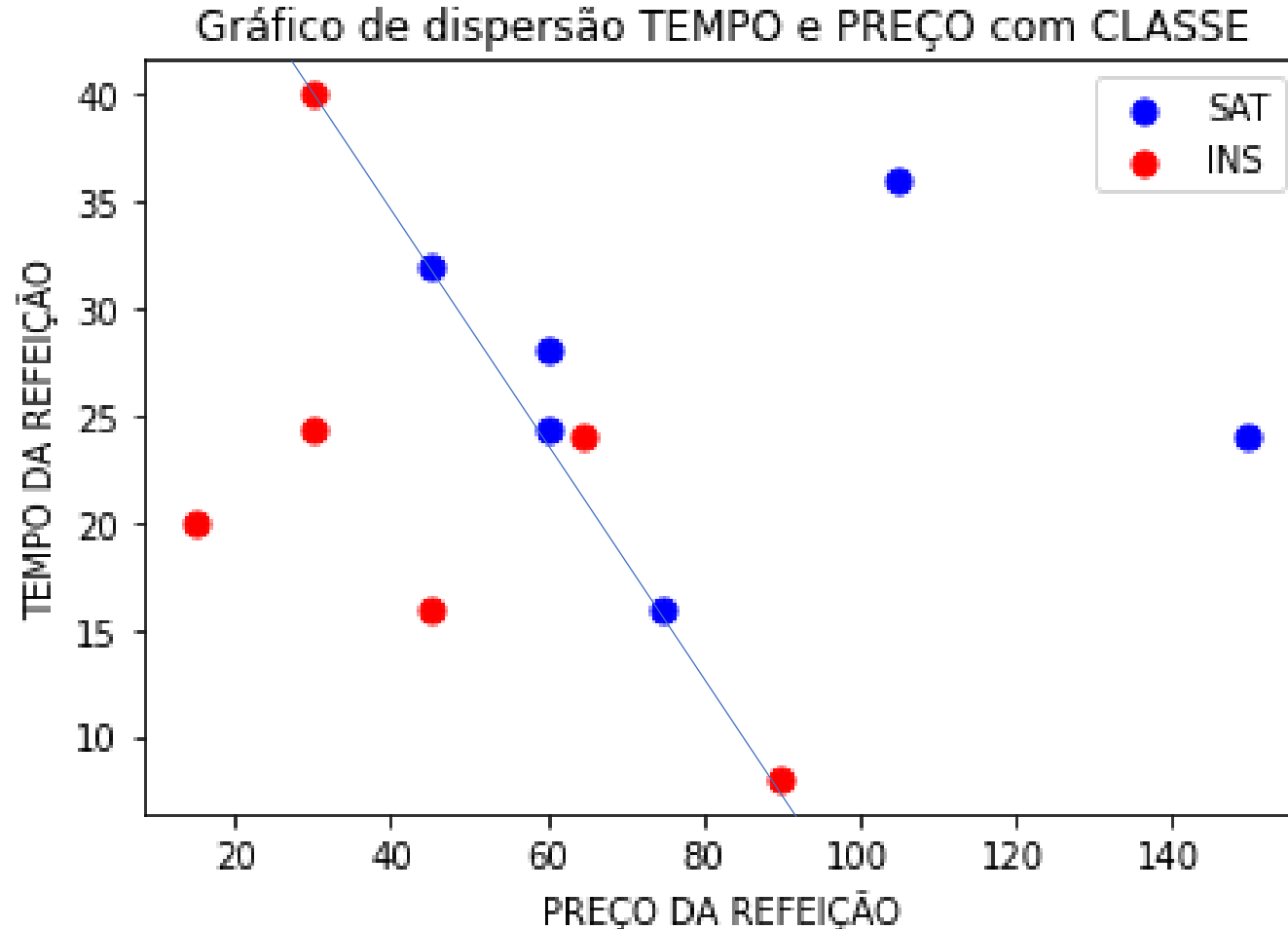


Gráfico de dispersão com cores por categoria

- Um gráfico de dispersão (*scatter plot*), também pode ser usado para visualizar como as categorias estão distribuídas
- No exemplo ao lado estão presentes os atributos
- PREÇO no eixo x
- TEMPO no eixo y
- CLASSE nas cores onde os clientes satisfeitos estão na cor azul e insatisfeitos na cor vermelha



Como é possível prever a satisfação dos clientes com base no tempo de preparo de uma refeição e seu preços ?

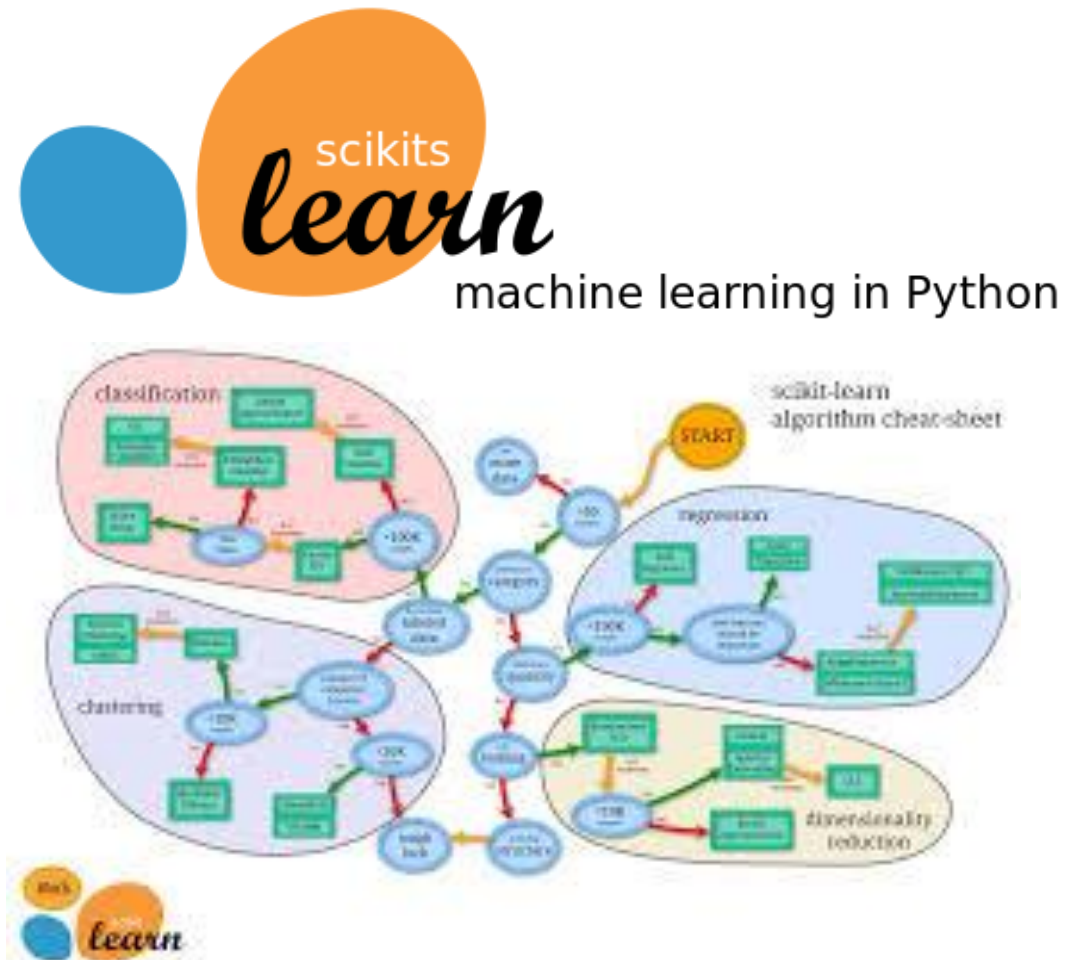


O processo de Classificação

- Os métodos gerais que os modelos de tipo sklearn implementam são:
- Analisar os dados
- Ajustar um modelo
- Prever classes
- Obtenha precisão

Aprendizado de Máquina com o Scikit Learn

- **Scikit-learn**
- A scikit-learn (originalmente scikits.learn) é uma biblioteca de aprendizado de máquina para a linguagem de programação Python.
- Ela inclui vários algoritmos de classificação, regressão e agrupamento incluindo máquinas de vetores de suporte, florestas aleatórias, gradient boosting, k-means e DBSCAN, e é projetada para interagir com as bibliotecas Python numéricas e científicas NumPy e SciPy.
- O projeto scikit-learn começou como scikits.learn, um projeto do Google Summer of Code por David Cournapeau.
- O nome resulta da ideia de que ele é um "SciKit" (um SciPy Toolkit), uma extensão do SciPy desenvolvida por terceiros e distribuída separadamente.[4]
- O código base original foi posteriormente reescrito por outros desenvolvedores. Entre os vários scikits, o scikit-learn (tal como o scikit-image) foi descrito como "bem mantido e popular" em novembro de 2012.
- Em 2015, scikit-learn está sob desenvolvimento ativo e é patrocinado por INRIA, Telecom ParisTech e ocasionalmente pelo Google (através do Google Summer of Code).[6]



Aprendizado de Máquina com o Scikit Learn

- A classificação é um mecanismo de aprendizado supervisionado para rotular uma amostra com base nos recursos.
- Aprendizado supervisionado significa que temos rótulos para classificação ou números para regressão que o algoritmo deve aprender.
- O Sklearn implementa muitos modelos comuns e úteis.
- Também veremos alguns que não estão no sklearn, mas implementamos a interface do sklearn.
- Como eles seguem a mesma interface, é fácil experimentar diferentes famílias de modelos e ver o desempenho delas.

Árvores de Decisão

- Uma árvore de decisão é como ir a um médico que faz uma série de perguntas para determinar a causa dos seus sintomas.
- Podemos usar um processo para criar uma árvore de decisão e ter uma série de perguntas para prever uma classe-alvo.
- As vantagens deste modelo incluem:
 - suporte para dados não numéricos (em algumas implementações),
 - pouca preparação de dados (sem necessidade de dimensionamento),
 - suporte para lidar com relacionamentos não-lineares,
 - são reveladas importâncias de recursos e é fácil de explicar.
- O algoritmo padrão usado para criação é chamado de árvore de classificação e regressão (CART).
- Ele usa a impureza de Gini (uma medida de entropia) ou a medida do índice para construir decisões.
- Isso é feito repetindo os atributos e encontrando o valor que oferece a menor probabilidade de classificação incorreta.

Passo 1: Criando a Matriz de Dados

- O primeiro passo é determinar de quais dados precisamos para responder à pergunta:
- Como é possível prever a satisfação dos clientes com base no tempo de preparo de uma refeição e seu preço ?
- Portanto se desejamos prever a satisfação dos cliente vamos considerar que CLASSE É A NOSSA VARÍAVEL ALVO
- Desejamos fazer essa predição com base no tempo de preparo de uma refeição e seu preço, portanto TEMPO E PREÇO SÃO OS ATRIBUTOS PREDITORES.

Passo 2 – Criar as Matrizes X e y

	y	X	
	CLASSE	PREÇO	TEMPO
0	SAT	60.0	28.0
1	SAT	75.0	16.0
2	SAT	60.0	24.4
3	SAT	105.0	36.0
4	SAT	45.0	32.0
5	SAT	150.0	24.0
6	INS	90.0	8.0
7	INS	45.0	16.0
8	INS	30.0	24.4
9	INS	64.5	24.0
10	INS	15.0	20.0
11	INS	30.0	40.0

Passo 3 – Separar as Matrizes X e y em train e test

	y	X		
	CLASSE	PREÇO	TEMPO	
0	SAT	60.0	28.0	Train 60%
1	SAT	75.0	16.0	
2	SAT	60.0	24.4	
3	SAT	105.0	36.0	
4	SAT	45.0	32.0	
5	SAT	150.0	24.0	
6	INS	90.0	8.0	Test 40%
7	INS	45.0	16.0	
8	INS	30.0	24.4	
9	INS	64.5	24.0	
10	INS	15.0	20.0	
11	INS	30.0	40.0	

O Fluxo da Modelagem: Ajuste, Predição e Precisão

X[:,0]	X[:,1]	y
60.0	28.0	SAT
75.0	16.0	SAT
60.0	24.4	SAT
105.0	36.0	SAT
45.0	32.0	SAT
150.0	24.0	SAT
90.0	8.0	INS
45.0	16.0	INS
30.0	24.4	INS
64.5	24.0	INS
15.0	20.0	INS

Criação do Modelo
`Model = DecisionTreeClassifier()`

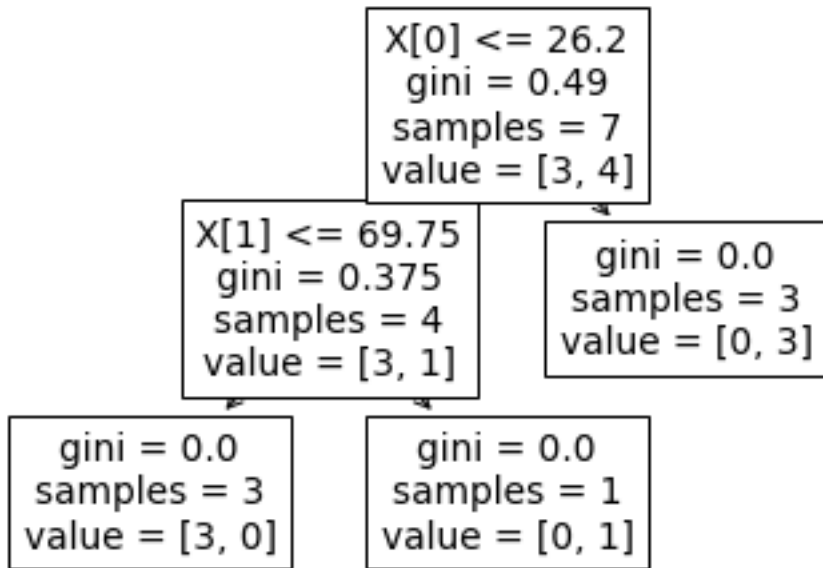
Ajuste do Modelo com os dados de
Treinamento
`Model.fit(X_train, y_train)`

Predição com os dados de
Treinamento
`y_pred = Model.predict(X_test)`

Avaliação da precisão com score
`Model.score(X_test, y_test)`

Avaliação do modelo com a matriz
de confusão
`Confusion_matrix(y_test, y_pred)`

Visualização da Arvore score = 0.4



VP = 1	FP = 2
FN = 1	VN = 1

Quizz

- Disponível em:
- <https://forms.gle/QYy2EQgAKds4When9>