
Previsões de venda usando anúncios de redes sociais

— Glevson da Silva Pinto —
Priscilla Amarante de Lima

Introdução

Estamos usando o conjunto de dados de anúncios da rede social e verificando a conversão de clique em compra com base no perfil do cliente.

Nossa base de dados pode ser acessada no Kaggle, em <https://www.kaggle.com/rakeshrau/social-network-ads> contendo um conjunto de dados categóricos para determinar se um usuário vai comprar um produto específico.

Objetivo: Descobrir se um usuário compra um produto clicando no anúncio no site com base em seu salário, idade e sexo.

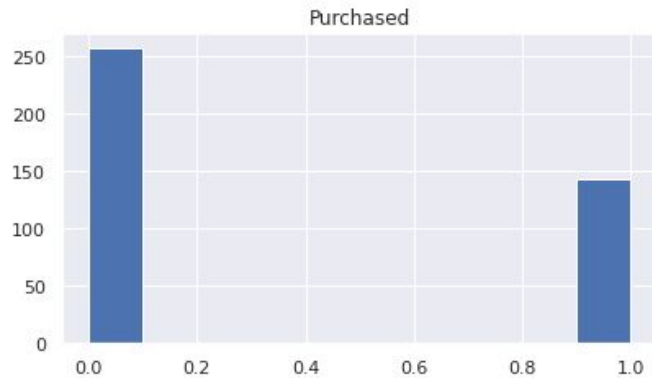
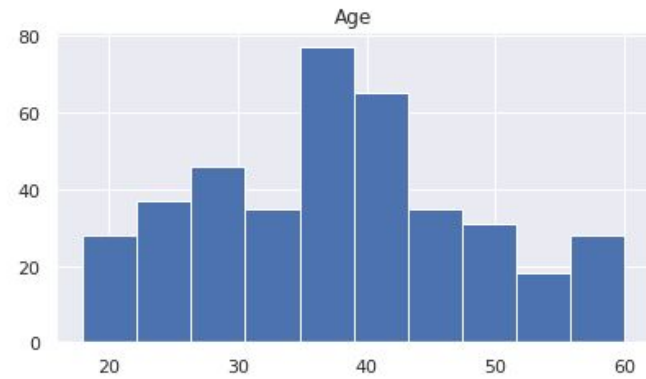
Estatística Descritiva

Dataset Social Networking

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
5	15728773	Male	27	58000	0
6	15598044	Female	27	84000	0
7	15694829	Female	32	150000	1
8	15600575	Male	25	33000	0
9	15727311	Female	35	65000	0

	User ID	Age	EstimatedSalary	Purchased
count	4.000000e+02	400.000000	400.000000	400.000000
mean	1.569154e+07	37.655000	69742.500000	0.357500
std	7.165832e+04	10.482877	34096.960282	0.479864
min	1.556669e+07	18.000000	15000.000000	0.000000
25%	1.562676e+07	29.750000	43000.000000	0.000000
50%	1.569434e+07	37.000000	70000.000000	0.000000
75%	1.575036e+07	46.000000	88000.000000	1.000000
max	1.581524e+07	60.000000	150000.000000	1.000000

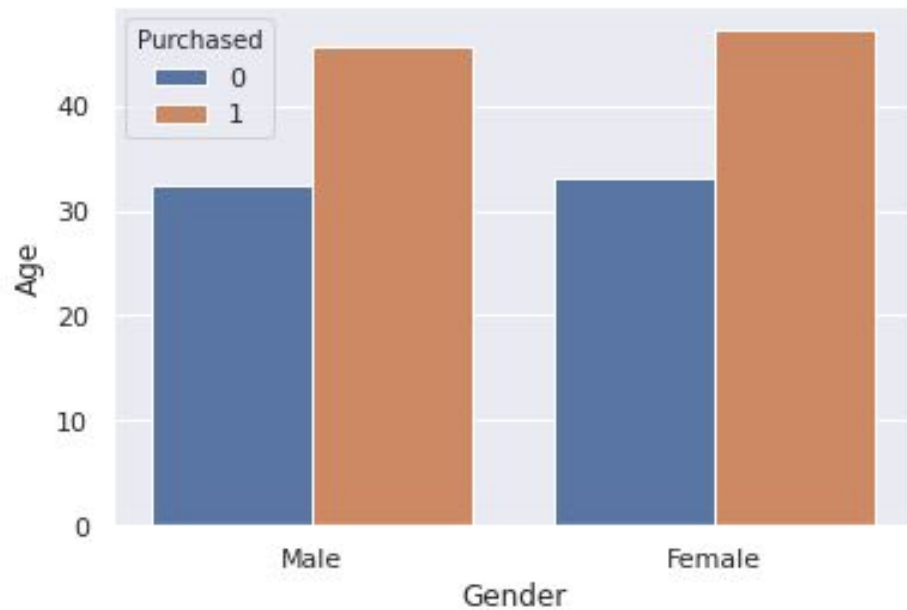
Estadística Descriptiva



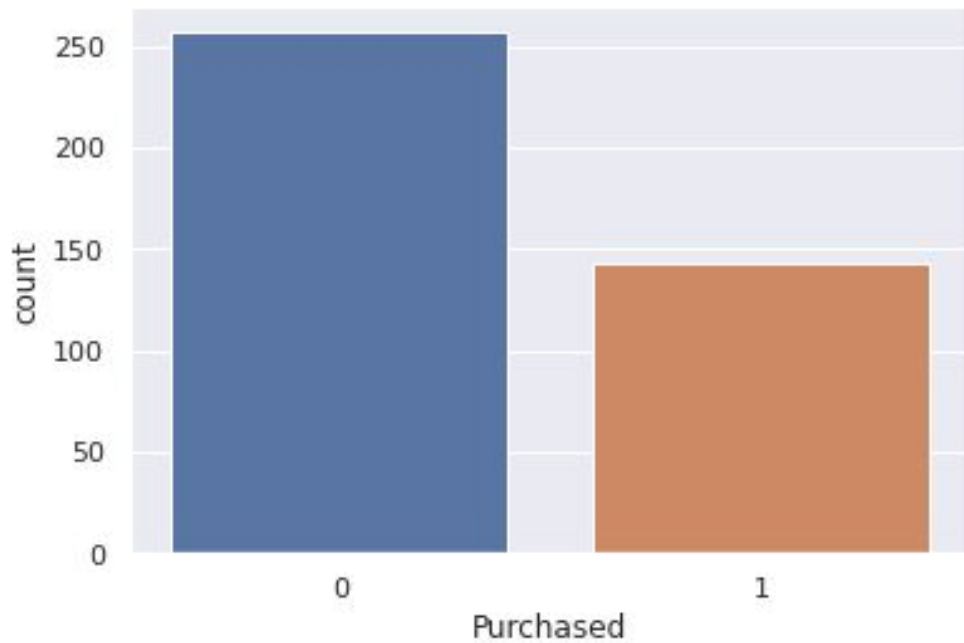
Correlação



Porcentagem de compradores por idade e sexo:



Proporção entre comprador(1) e não comprador (0)



Divisão do conjunto de dados: treinamento e teste

Estamos fornecendo o tamanho do teste como 0,20, o que significa que nossa amostra de treinamento contém 320 conjuntos de treinamento e a amostra de teste contém 80 conjuntos de teste.

Modelo Naive Bayes - GaussianNB

De modo geral, o Naive Bayes é usado para prever a chance de um evento com base nos dados existentes. Além disso, pelo fato de ele ser ingênuo ("ingênuo"), ele ignora a relação entre as variáveis, pois só temos números que representam texto, e a relação entre as variáveis é quase zero.

Usaremos o GaussianNB - Por se tratar de dados contínuos, uma suposição típica é que os valores contínuos associados a cada classe são distribuídos de acordo com uma distribuição normal (ou gaussiana).

No projeto assumimos features obedece a uma distribuição de probabilidade gaussiana.

```
1 #Treinamento do modelo Naive Bayes no conjunto de treinamento
2 from sklearn.naive_bayes import GaussianNB
3 classifier = GaussianNB()
4 classifier.fit(X_train, y_train)
```

Resultado dos testes

Para os primeiros 8 valores, ambos são iguais.

Podemos avaliar usando a matriz de confusão e a pontuação de precisão, comparando os valores de teste previstos e reais

```
1 y_pred
```

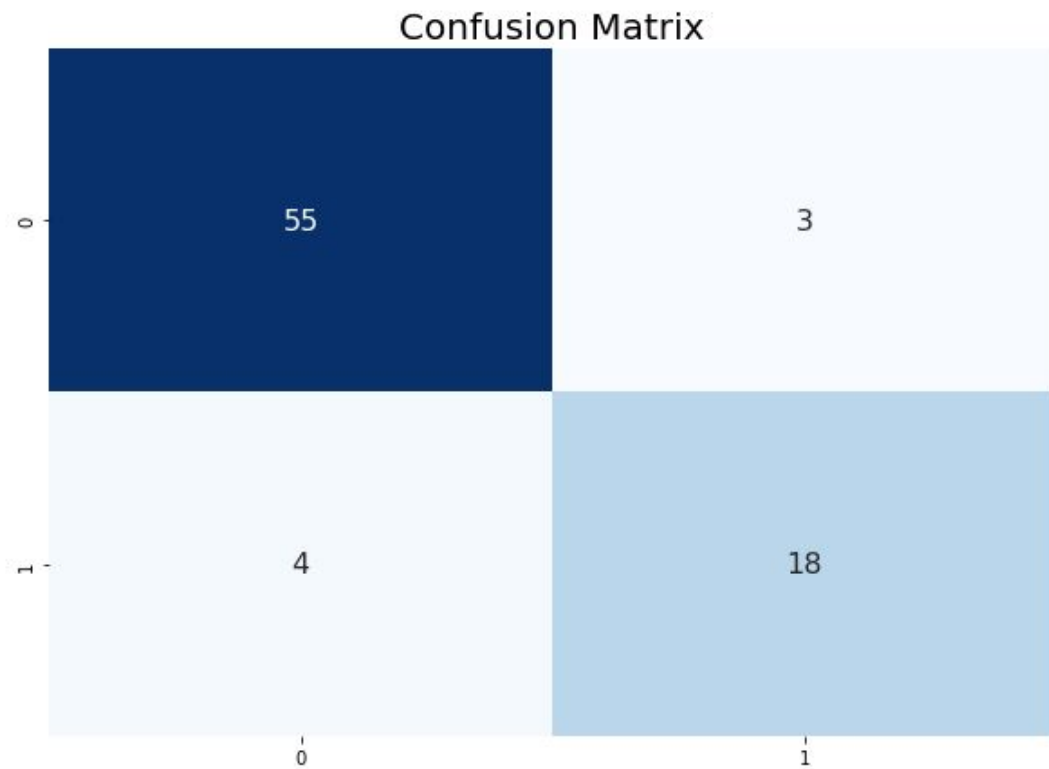
```
array([0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1,  
       0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,  
       1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1,  
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1])
```

```
1 y_test
```

```
array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1,  
       0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,  
       1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1,  
       0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1])
```

Matriz de confusão

Acurácia= 0.9125



Usando Cross Validation K = 5

O Cross Validation envolve a divisão dos dados em vários conjuntos (partes), um dos quais é usado para treinamento e o outro é usado para testar e avaliar o desempenho do modelo.

```
1 from sklearn.model_selection import cross_val_score
2 pred_gnb = cross_val_score(GaussianNB(), X, y, cv=5)
```

```
1 pred_gnb
```

```
array([0.8125, 0.9625, 0.925 , 0.8125, 0.9   ])
```

Conclusão

- Os resultados obtidos utilizando as técnicas estatísticas foram:
A faixa etária que mais clica em anúncios em sites estão entre 25 - 45 anos;
Acurácia satisfatória acima de 90%

Obrigado!!