

Classificação de venda

Modelos utilizados

- Random Forest
- Regressão Logística
- KNN
- Naïve Bayes Gaussiano
- Árvore de Decisão
- XGBoost

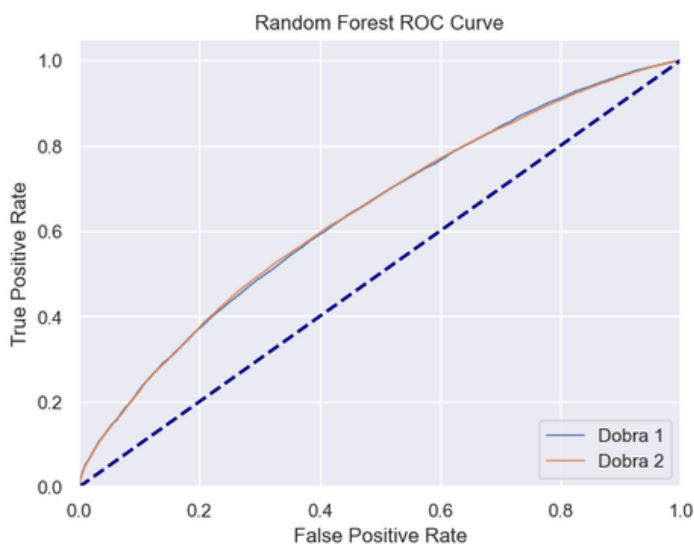
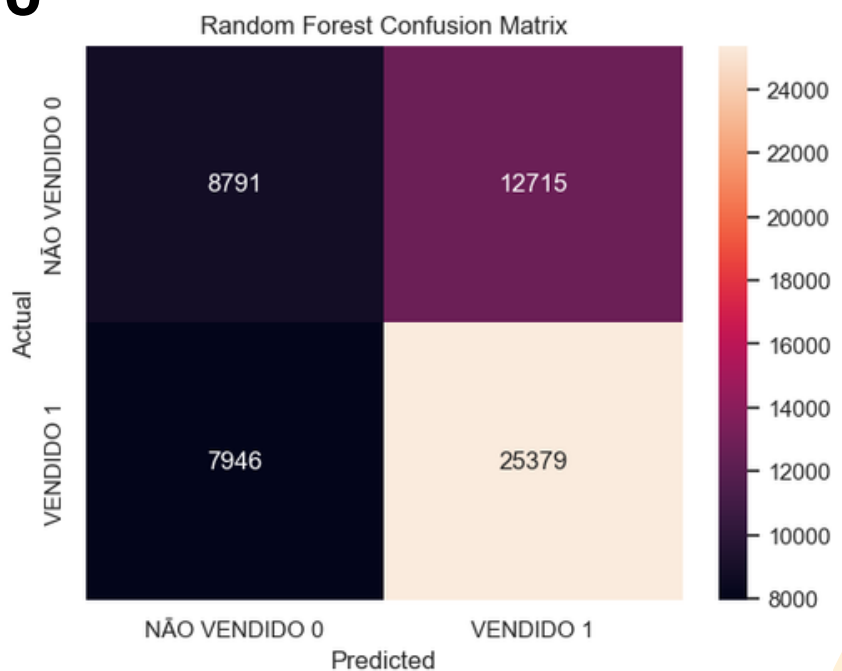
Intenção do Modelo

Os modelos foram treinados para classificar se um item de roupa usada que entra em um brechó será vendido ou não.

Métricas de Avaliação

Random Forest

- Acurácia: 0,63
- Precisão (classe 0): 0,53
- Precisão (classe 1): 0,67
- Recall (classe 0): 0,42
- Recall (classe 1): 0,76
- F1-score (classe 0): 0,47
- F1-score (classe 1): 0,71



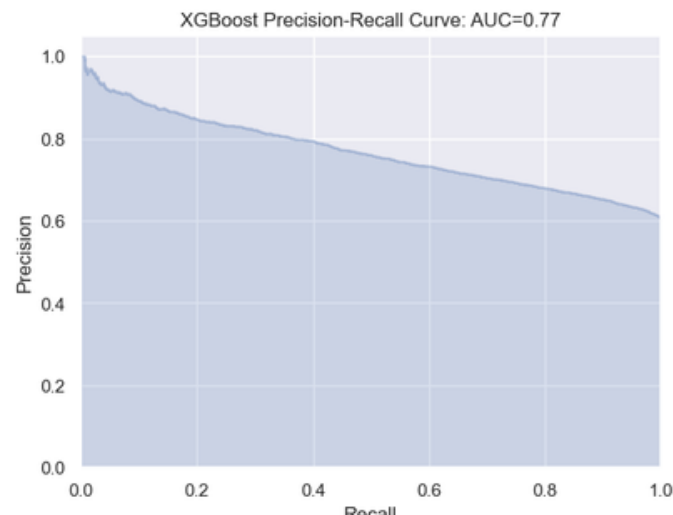
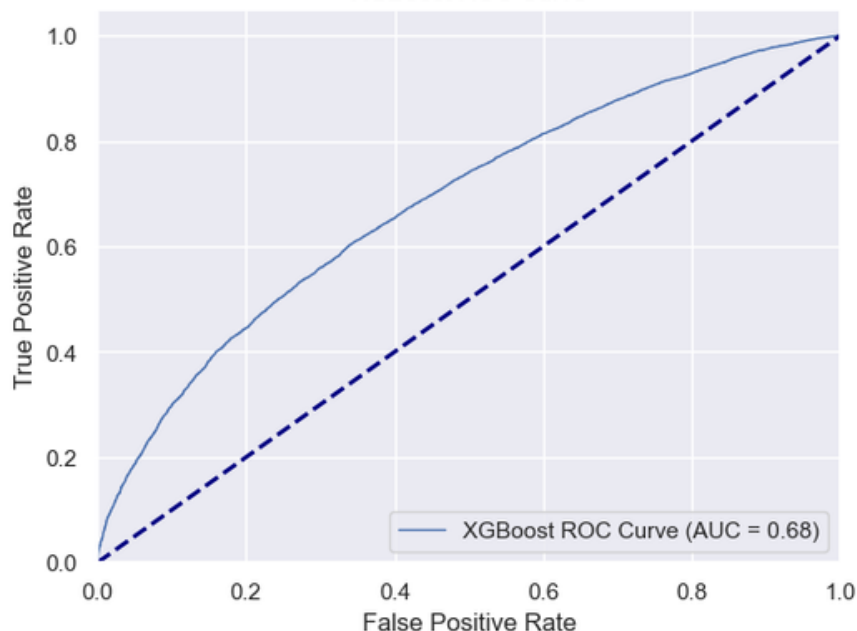
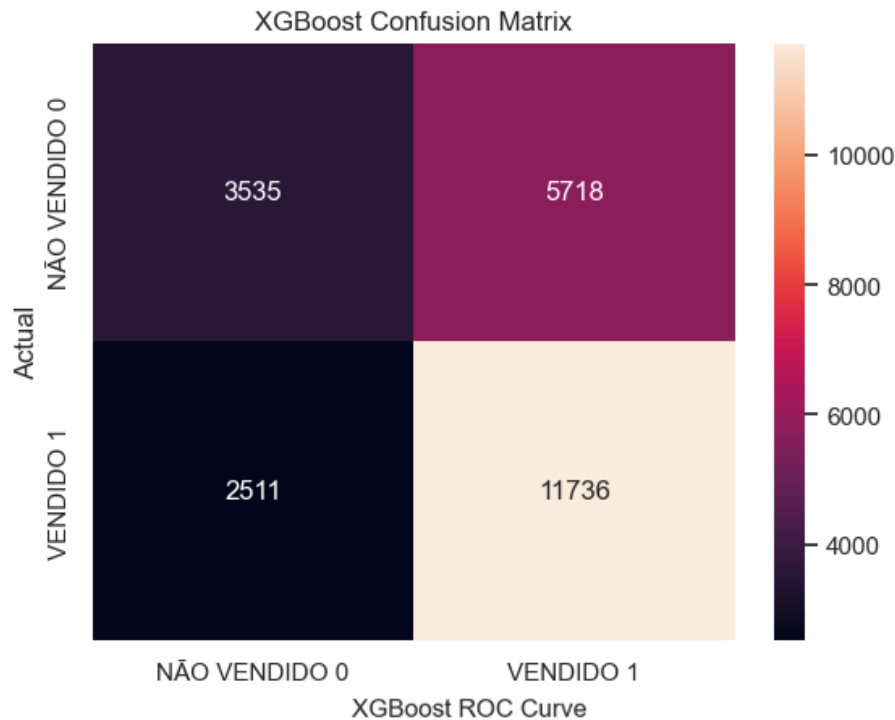
XGBoost

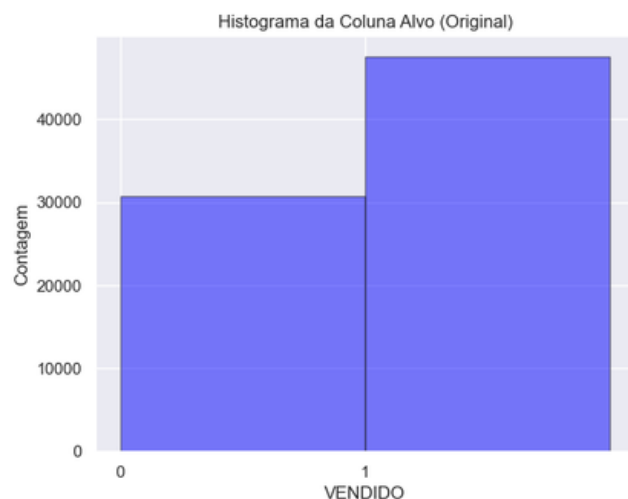
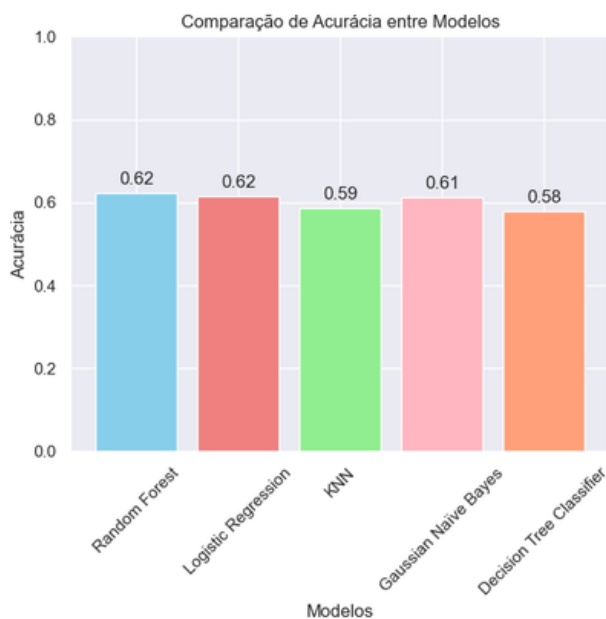
- Acurácia: 0,65
- Precisão (classe 0): 0,58
- Precisão (classe 1): 0,67
- Recall (classe 0): 0,38
- Recall (classe 1): 0,82
- F1-score (classe 0): 0,46
- F1-score (classe 1): 0,74

Pré-processamento de Dados

Foi utilizado principalmente funções feitas a mão para categorizar todos os dados das colunas, já que se portavam em formas de classes, um exemplo é a coluna Tamanho, que foi mudada para 5 classes apenas e normalizada em pp, p, m, g, gg.

Depois de normalizadas utilizamos o LabelEncoder, que atribui um número a uma string, ou seja agora todas novas variáveis eram numéricas mesmo sendo categóricas





Conclusão

A base de dados era extensiva e complexa de se mexer, uma vez que vários valores eram categóricos e sem padronização, principalmente a coluna de marcas, que se fosse normalizada com certeza melhoraria bastante a performance dos modelos. Uma das muitas limitações ao longo do treinamento foi a falta de processamento computacional, gostaríamos de ter feito 10 validações cruzadas ao invés de 2, e mais de 100 epochs nos treinos da rede neural, mas o resultado foi bastante satisfatório.

	marca	precoComDesconto	precoSemDesconto	Condicao	VENDIDO	tamanho_mapeado	Cores_Agrupadas
0	8601	64.00	160.0	0	0	2	21
1	10361	57.00	219.0	0	1	3	16
2	18	69.00	279.0	0	0	3	1
3	2160	25000.00	87500.0	4	1	3	18
4	7807	20.97	69.9	0	0	2	14
...
78326	6062	15.00	69.9	0	0	2	18
78327	6656	33.75	135.0	0	1	3	3
78328	5802	27.48	89.9	0	1	1	14
78329	10361	7000.00	28000.0	4	1	3	2
78330	10361	4000.00	15000.0	4	1	3	16