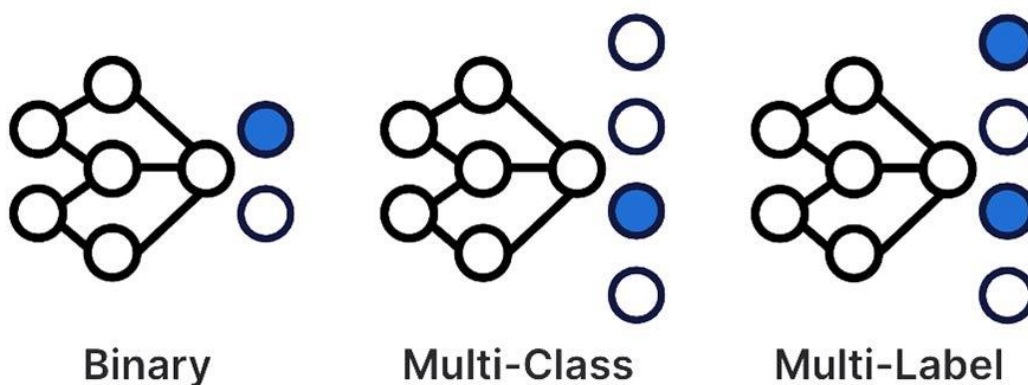


Relatório da análise de dados

Programa Lighthouse -Indicium

Guilherme Frutuoso de Almeida

Tendo em vista o formato dos dados a descrição dos requisitos, fica claro que a tarefa preditiva é de classificação. Analisando a coluna “Label”, podemos especificar ainda mais essa tarefa, chamando-a de “classificação de múltiplos rótulos”.



Tarefas assim precisam de algoritmos específicos que se encaixem com ela e com seu tipo de estrutura de dados. Operações vetoriais (Em “Deep Learning”, Tensor Operations”) são um dos motores da aprendizagem de máquina e dependendo da tarefa e de seu conjunto de dados, a estrutura(Vetores e Matrizes) dos subconjuntos de dados podem mudar, o que afeta diretamente os modelos.

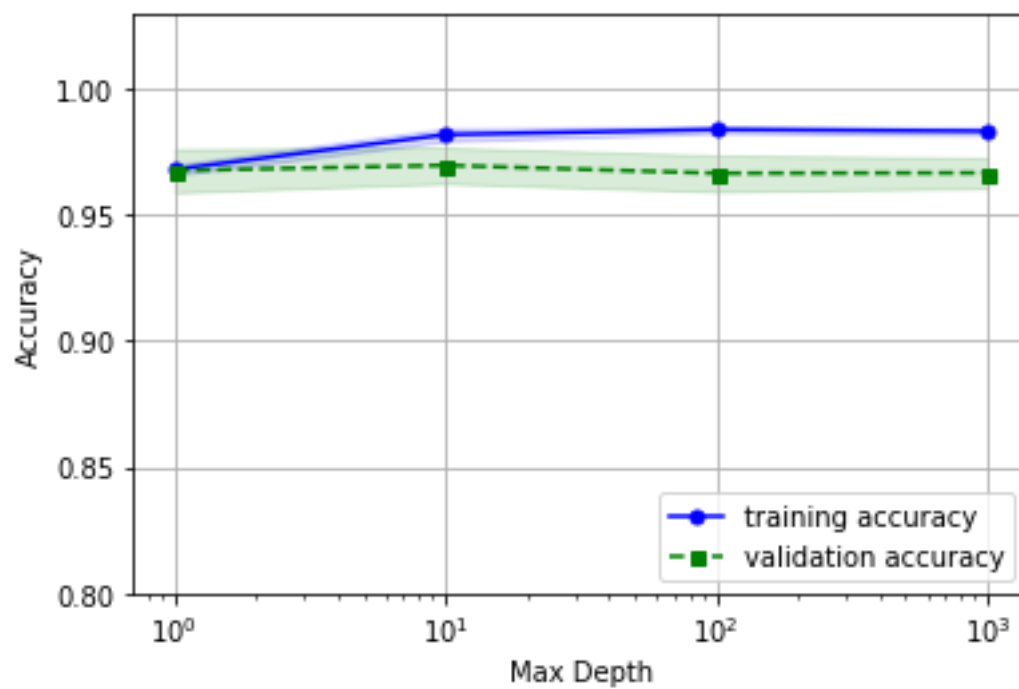
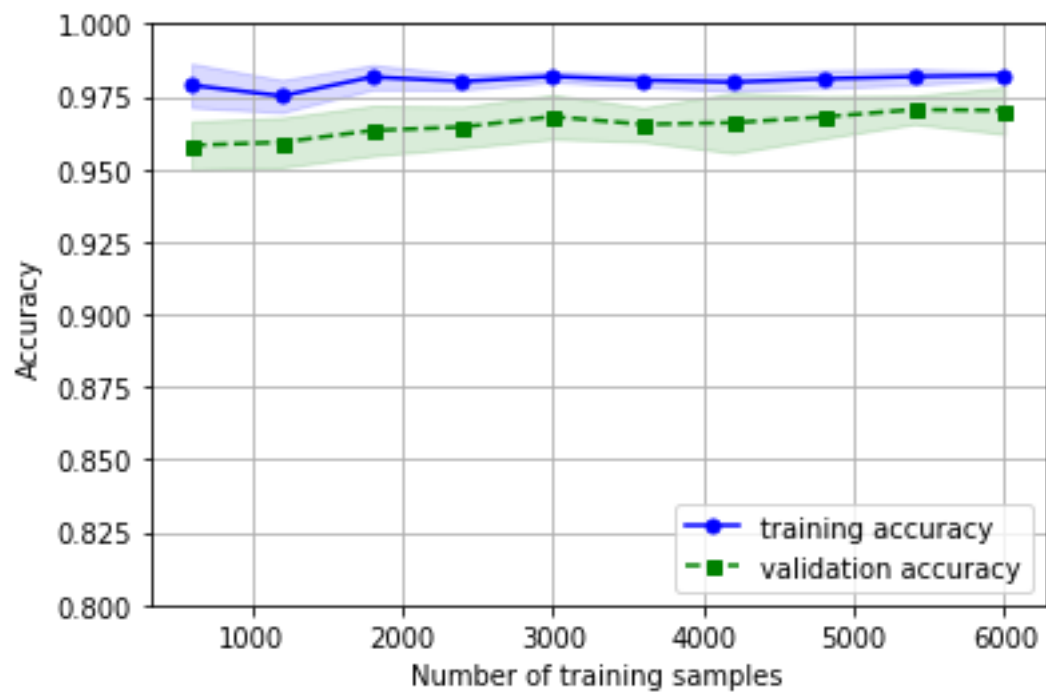
Portanto, “jogando na bola de segurança” e ao mesmo tempo, “ousando mais”, decidi por dois modelos: Árvore de Decisão(01) e o ensemble model Random Forest em conjunto com o XGBoost(02). O primeiro tem parametrização simples, o outro mais complexa.

RESULTADOS:

Modelo 01

-Acurácia: 96.81962744207179

-Erro quadrado Médio: 0.009844010298349236



Modelo 02

XGBoost Accuracy: 0.9879518072289156

XGBoost F1-Score (Micro):

