Yuri Carlos Bonifácio Neves **1609670**

Árvore de Decisão

Por meio do trabalho é possível perceber que o método da árvore de decisão é realmente muito sensível aos parâmetros como número máximo de folhas ou mesmo a profundidade da própria árvore. Rodando a árvore para valores maiores de profundidade, como por exemplo uma profundidade de 10, pude obter uma taxa que variava de 76 até 79 de accuracy. Limitando essa profundidade a valores próximos a 5 já pude perceber grandes melhoras na árvore (acredito que isso se deva porque a árvore acaba sendo mais genérica no fim de suas decisões).

É interessante também ser cuidadoso na seleção das características e no tipo de dados que se coloca na aprendizagem por meio da árvore. Percebi que para dados mais categóricos pode-se obter valores mais interessantes. Para o vetor de características [sex, embarked, pclass , family, title] pude obter um valor de 0.85 mais ou menos.

Para dados não categóricos, ou seja, com um range maior de possibilidades, como o Fare ou o Sib e outros vetores temos também valores bons, porém acho que isso se deve a um possível overfitting, já que dados desse tipo tendem as separar muito os dados, deixando as folhas da árvore muito específicas.

Ao fim do processo, executando o vetor de características foi [sex, embarked, pclass , family, title], com o grid_search escolhendo parâmetros, em geral, dessa maneira:
{'min_samples_split': 10, 'max_leaf_nodes': 20, 'max_depth': None, 'min_samples_leaf': 1}

Naive Bayes

Nesse método, variando com dados não categóricos, pro vetor de características [sex, imputed, sib, ticket] (fora sex, são não categóricos) obtive uma taxa de 0.46. Já para dados categóricos obtive 0.84 (isso para o vetor [sex, embarked, pclass , family, title]). Podemos ver bem a diferença nos resultados modificando somente o tipo dos dados.