Prediction Results

Classification (decision tree):

- Our dataset is about movies and it has a lot of numeric data, so we used decision tree algorithm in our classification and prediction process.
- The Decision Tree Selects the best attribute (most informative) using Attribute Selection Measures like (Information Gain) to split the records.
- That attribute will be a decision node and breaks the dataset into smaller subsets.
- Then recursively repeat that process for each child node until we satisfy our classification process by running out of attributes or instances.
- We are trying to predict which IMDB Category each given movie is in.
- The IMDB Categories are 3 classes the first is movies that its imdb score is higher than or equal 7.8, the second class is movies with score between 7.8 and 5.5 and the third class is movies with score under 5.5.(IMDB is a movies website that collects professional critics and audiences' movies reviews, and each review contains a score for the movie on a scale from 0.0 to 10.0 the imdb_score attribute is the average of all reviews scores).
- We used scikit's library DecisionTreeClassifier to build our model and split our dataset 70% for training and 30% for testing.
- We used attributes like gross, budget, title year, number of Facebook likes and number of voted users as our input data to the algorithm.
- After our model learns from training data from the data set its ready for evaluation so we computed the Accuracy by comparing actual test set values and predicted values.
- The accuracy of the model is approximately 74%.