Name : Meer murtaza

Project report

So first I load the dataset and apply some data exploration operations like see top five values check the size of data set describe dataset and check the correlation between columns apply data visualization

Operations using histogram and count plot to analyze the data box plot to check outliers I also check missing values but there isn’t check the duplicates check the target column using count plot and after that use correlation metrics to identify the relation between columns some of them have weak relation with my target column so I drop that columns (Age, Employee\_count …) after that I convert string column into integer as I can apply the model after that I apply models RandomForest accuracy = 85%

KNN accuracy = 83% , GradientBoostingClassifier accuracy = 87% DecisionTreeClassifier accuracy = 77%

Gaussian Naive Bayes accuracy = 84%, LogisticRegression accuracy = 85%