**Machine Learning**

**Predicting Fake Job Postings**

**Name: AMIT KUMAR SAHU**

**Reg.No: 18MIS7250**

**Objective**

To predict if a job posted is fake or not.

|  |  |  |
| --- | --- | --- |
| **Dataset Link** | **Number of rows and Columns** | **About columns** |
| <https://www.kaggle.com/shivamb/real-or-fake-fake-jobposting-prediction> | Rows: **17880**  Columns: **18** | **Number of Categorical columns: 13**  **Number of Integer/Float Columns: 5** |

|  |  |
| --- | --- |
| **Algorithms Applied with Label Encoding** | **Performance(Accuracy/other confusion matrix measures)** |
| **Algorithm 1-( Logistic Regression )** | accuracy score: 0.9530  Confusion Matrix:  [[1984 5]  [ 93 2]]  ROC AUC: 0.5093 |
| **Algorithm 2-( Naive Bayes )** | accuracy score: 0.9309  Confusion Matrix:  [[1919 70]  [ 74 21]]  ROC AUC: 0.5929 |
| **Algorithm 3-( Decision Trees )** | accuracy score: 0.9616  Confusion Matrix:  [[1953 36]  [ 44 51]]  ROC AUC: 0.7594 |
| **Algorithm 4-( Random Forest )**  **250 DTs** | accuracy score: 0.9765  Confusion Matrix:  [[1987 2]  [ 47 48]]  ROC AUC: 0.7521 |
| **Algorithm 5-( Random Forest )**  **1000 DTs** | accuracy score: 0.9765  Confusion Matrix:  [[1986 3]  [ 46 49]]  ROC AUC: 0.7571 |
| **Algorithm 6-( KNN )**  **K=5** | accuracy score: 0.9592  Confusion Matrix:  [[1971 18]  [ 67 28]]  ROC AUC: 0.6428 |
| **Algorithm 7-( KNN )**  **K=6** | accuracy score: 0.9611  Confusion Matrix:  [[1981 8]  [ 73 22]]  ROC AUC: 0.6138 |
| **Algorithm 7-( SVM )**  **Kernel = RBF** | accuracy score: 0.9544  Confusion Matrix:  [[1989 0]  [ 95 0]]  ROC AUC: 0.5000 |
| **Algorithm 8-( Bagging )**  **Base estimater: SVM** | accuracy score: 0.9544  Confusion Matrix:  [[1989 0]  [ 95 0]]  ROC AUC: 0.5000 |
| **Algorithm 9-( Gradient Boosting )** | accuracy score: 0.9731  Confusion Matrix:  [[1982 7]  [ 49 46]]  ROC AUC: 0.7403 |
| **Algorithm 10-( XG Boost )** | accuracy score: 0.9683  Confusion Matrix:  [[1983 6]  [ 60 35]]  ROC AUC: 0.6827 |
| **Algorithm 11-( Stacking -> DT & RF)** | Ensemble ROC-AUC score: 0.943 |
| **Algorithm 12-( Stacking -> Decision trees, Random forest & Xg boost)** | accuracy score: 0.9755  Confusion Matrix:  [[1980 9]  [ 42 53]]  ROC AUC: 0.7767 |
| **Algorithm 13 – ANN**  **Input activation function: relu**  **Hidden layer [128] activation function: relu**  **Output activation function: sigmoid**  **Batch size: 10**  **Epochs: 10** | Accuracy score: 0.9611253196930947  Confusion Matrix:  [[1866, 6],  [ 70, 13]]  ROC-AUC score: 0.5767106889094841 |
| **Algorithm 13 – ANN**  **Input activation function: relu**  **Hidden layer [128] activation function: relu**  **Output activation function: sigmoid**  **Batch size: 10**  **Epochs: 100** | Accuracy score: 0.9677749360613811  ROC-AUC score: 0.781677994027391 |