

Data Mining & Data Warehousing Project

Real / Fake Job Posting Prediction Final Report

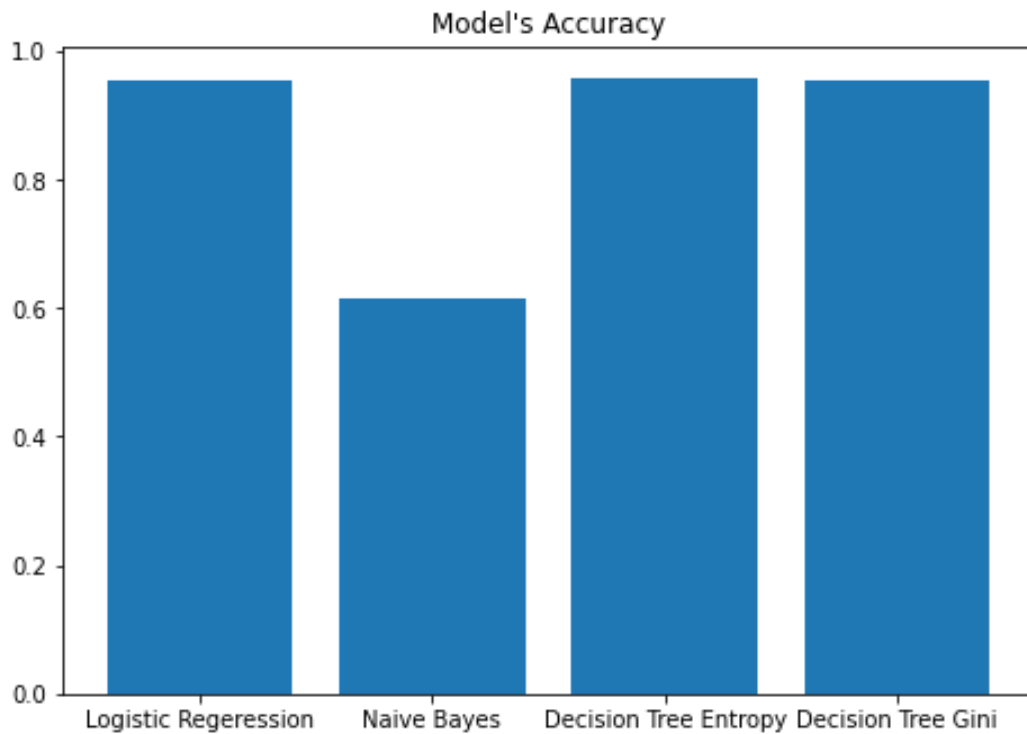
Name: **Muhammad Hazique Khatri**

Seat No: **B18101071**

Task	Description
Source	Link: https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction?resource=download&select=fake_job_postings.csv
Columns & Rows Count	18, 17880
Pre Processing	<ul style="list-style-type: none">Deleted unnecessary columns.Converted all column text in single text format.Preprocessed using nltk library.Used word_tokenize for creating tokens.Used word_lemmatize for useful words in English.Used CountVectorizer to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text.
Training Data	80%
Testing Data	20%
Model Evaluation	<ol style="list-style-type: none">Logistic Regression<ul style="list-style-type: none">Accuracy Value = 0.9555369127516778AUC Value = 0.8934Naive Bayes<ul style="list-style-type: none">Accuracy Value = 0.6160514541387024AUC Value = 0.8269Decision Tree Entropy<ul style="list-style-type: none">Accuracy Value = 0.959731543624161AUC Value = 0.7853Decision Tree Gini<ul style="list-style-type: none">Accuracy Value = 0.9555369127516778AUC Value = 0.7778

GitHub Link: https://github.com/HaziqueIqbal/Real-Fake_Job_Posting_Prediction

Accuracy Graph



ROC Curve

