Decision Tree Model for Fake News Detection

Author: Robin Chen

Student ID:210178740

This report summarizes the results of using a Decision Tree Classifier to classify news headlines as real or fake. The dataset consists of labeled headlines extracted from real.txt and fake.txt.

Preprocessing: Headlines were vectorized using CountVectorizer.

Model Training: A DecisionTreeClassifier was trained with varying max_depth values.

Evaluation: Validation and test accuracies were measured.

The performance of the classifier at different depths is shown below:

Max Depth	Validation Accuracy	Test Accuracy
2	0.6224	—
5	0.6347	—
10	0.6592	—
20	0.6837	—
50	0.7184	0.7551

The decision tree model performs well with max_depth = 50 and captures the key patterns in the data. Further improvements may include using TF-

IDF vectorization, pruning techniques, or ensemble models such as random forests.