

Decision Tree Model for Fake News Detection

We classify news headlines as true or false using a decision tree classifier.

The dataset includes labeled headlines extracted from real.txt and fake.txt.

First we vectorize the headlines using CountVectorizer, then train a decision tree classifier with different max_depth values and finally measure validation and test accuracy.

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.