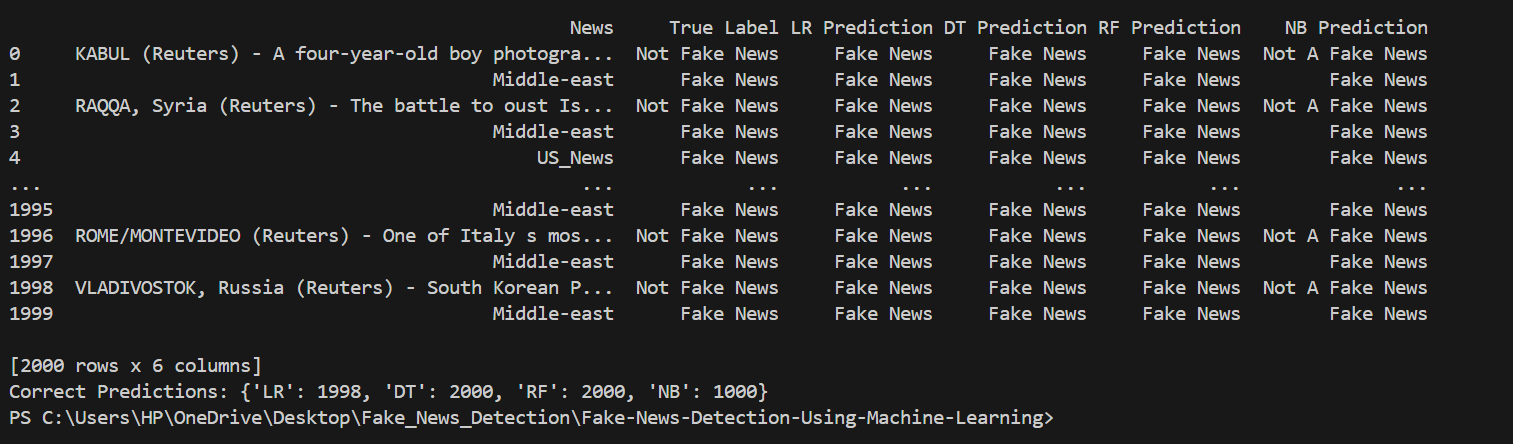
Fake News Detection Results

# Word Count Feature



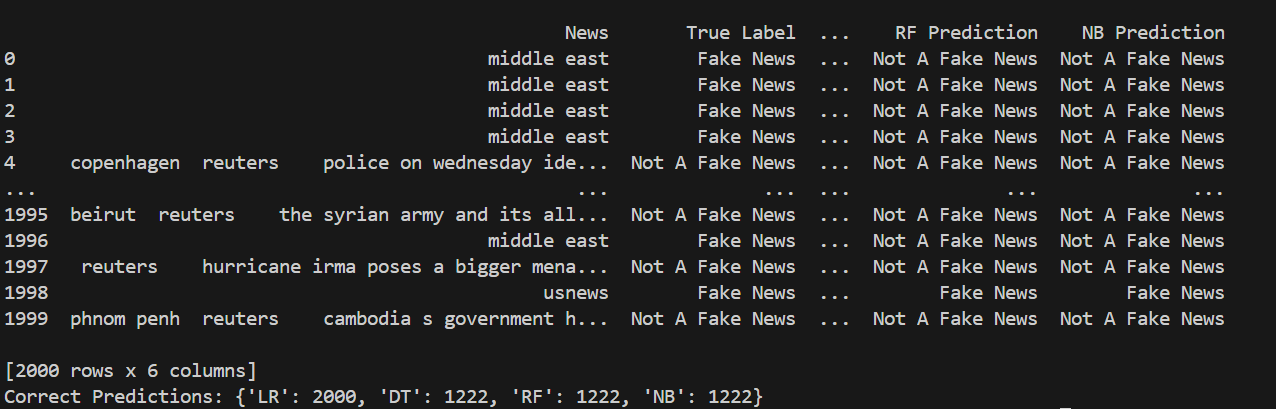
**LR = 1998/2000 Correct** , **99.9 % Accuracy**

**DT = 2000/2000 Correct , 100 % Accuracy**

**RF = 2000/2000 Correct , 100 % Accuracy**

**NB = 1000/2000 Correct , 50 % Accuracy**

# N-Gram Count Feature



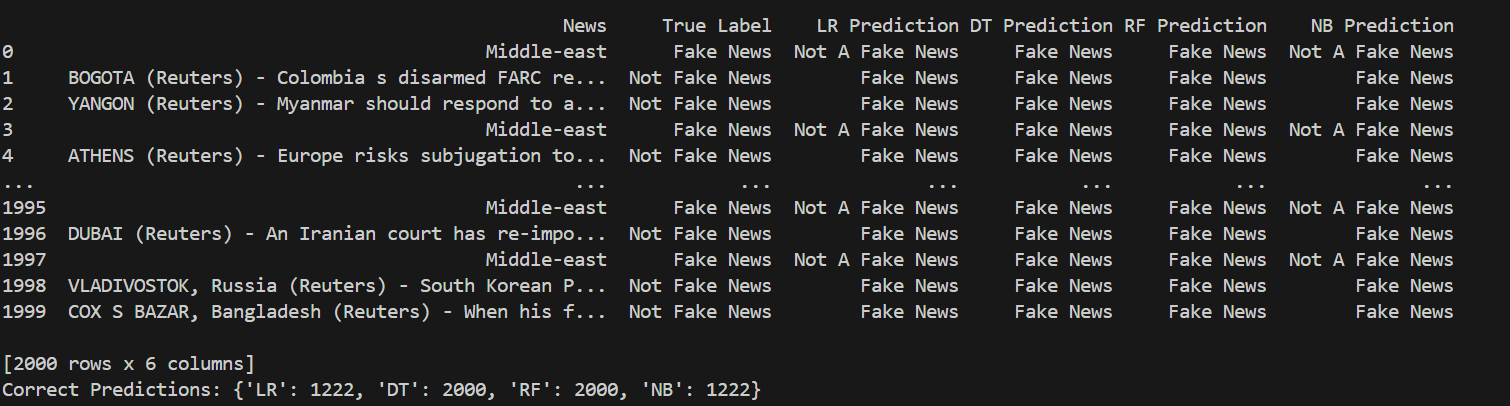
**LR = 2000/2000 Correct** , **100 % Accuracy**

**DT = 1222/2000 Correct , 61 % Accuracy**

**RF = 1222/2000 Correct , 61 % Accuracy**

**NB = 1222/2000 Correct , 61 % Accuracy**

# TF-IDF Feature



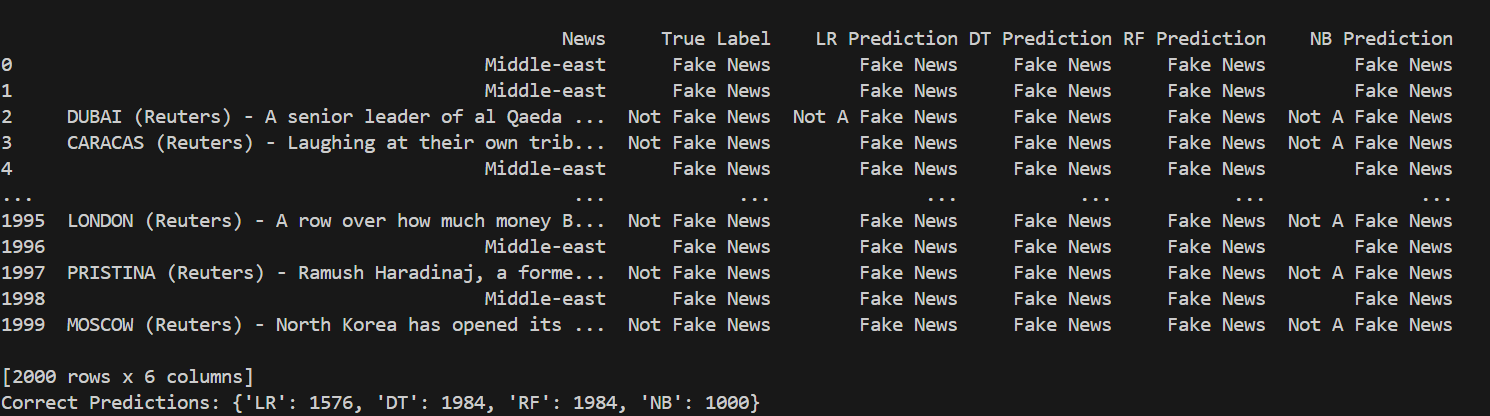
**LR = 1222/2000 Correct**, **61 % Accuracy**

**DT = 2000/2000 Correct, 100 % Accuracy**

**RF = 2000/2000 Correct, 100 % Accuracy**

**NB = 1222/2000 Correct ,61 % Accuracy**

# Sentiment Analysis Feature



**LR = 1576/2000 Correct** , **78.8 % Accuracy**

**DT = 1984/2000 Correct , 99.2 % Accuracy**

**RF = 1984/2000 Correct , 99.2% Accuracy**

**NB = 1000/2000 Correct , 50% Accuracy**

**Fake News Detection Project: Overview**

**Objective:**

**The primary aim of this project was to evaluate different feature extraction techniques and classifiers for the task of fake news detection. The goal was to determine which combination of feature and classifier works best in terms of accuracy and performance on this specific task.**

**Feature Extraction Techniques:**

1. **Word Count:** The number of words in each news article was used as a feature.
2. **N-Gram Count**: This feature considered the frequency of sequences of 'n' words (bigrams, trigrams, etc.).
3. **TF-IDF (Term Frequency-Inverse Document Frequency**): This feature measures how important a word is to a document in a collection.
4. **Sentiment Analysis:** This feature captures the overall sentiment (positive or negative) of the news article.

**Classifiers Used:**

1. **Logistic Regression (LR)**
2. **Decision Tree (DT)**
3. **Random Forest (RF)**
4. **Naive Bayes (NB)**

**Results Analysis:**

* **Word Count Feature**: Decision Tree and Random Forest classifiers performed exceptionally well, achieving 100% accuracy. Logistic Regression also performed strongly with nearly perfect accuracy. However, Naive Bayes struggled with this feature, achieving only 50% accuracy.
* **N-Gram Count Feature**: Logistic Regression was the standout performer, achieving perfect accuracy, while other classifiers, including Decision Tree, Random Forest, and Naive Bayes, reached only 61% accuracy.
* **TF-IDF Feature**: Decision Tree and Random Forest again excelled with 100% accuracy, but Logistic Regression and Naive Bayes lagged behind with only 61% accuracy.
* **Sentiment Analysis Feature**: Decision Tree and Random Forest were again top performers with 99.2% accuracy. Logistic Regression achieved 78.8% accuracy, while Naive Bayes remained at 50%.

**Conclusion:**

* **Best Features:** The Word Count and TF-IDF features, in combination with Decision Tree and Random Forest classifiers, provided the best results for fake news detection, with 100% accuracy in multiple tests**.**
* **Best Classifiers:** Decision Tree and Random Forest consistently outperformed other classifiers across different feature sets. Logistic Regression performed well with the N-Gram Count feature but showed lower accuracy with other features. Naive Bayes underperformed across most feature sets, achieving the highest accuracy (61%) only with N-Gram Count and TF-IDF features**.**

**Overall, Decision Tree and Random Forest classifiers with Word Count or TF-IDF features appear to be the most effective combination for this fake news detection task.**