

Fake News Detection

Mythbusters

Outline

1. Introduction (Recap)
2. System Architecture (Recap)
3. Data Source (Recap)
4. Data Preprocessing (Recap)
5. Exploratory Data Analysis (EDA)
6. Machine Learning Implementation
7. Result Evaluation
8. Application Design

Introduction

The fake news on social media and various other media is wide spreading and is a matter of serious concern due to its ability to cause a lot of social and national damage with destructive impacts. A lot of research is already focused on detecting it. In this project, we will implement Intelligent System to detect Fake News

System Architecture

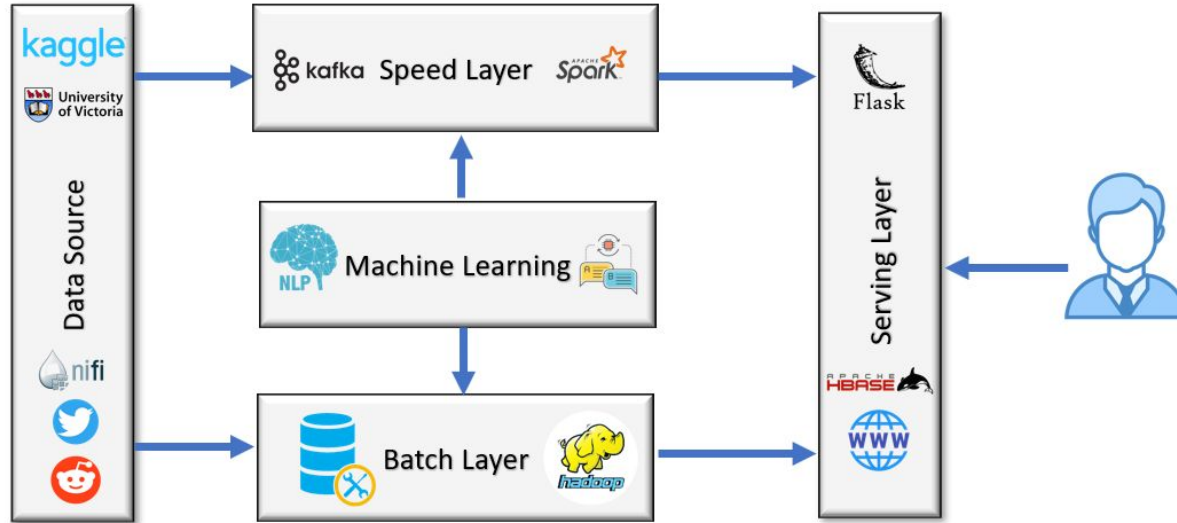


Figure 1: Lambda Architecture

Data sources

- Labeled batch data
 - Fakeddit → 1M reddit posts
 - LIAR → 12,8K social media statements
 - ISOT → 40K articles
- Streaming data
 - Twitter
 - Reddit

artificially streamed to a daily frequency

Data Preprocessing (PySpark) Con't

1. Text Cleaning

- a. Handle Missing Value
- b. Convert Text into Lowercase letter
- c. Remove Punctuation from Text
- d. Remove Urls from Text
- e. Remove @tags
- f. Remove Special Characters

2. Preprocess Operations

- a. Tokenization
- b. Removing Stop Words

Data Preprocessing (PySpark)

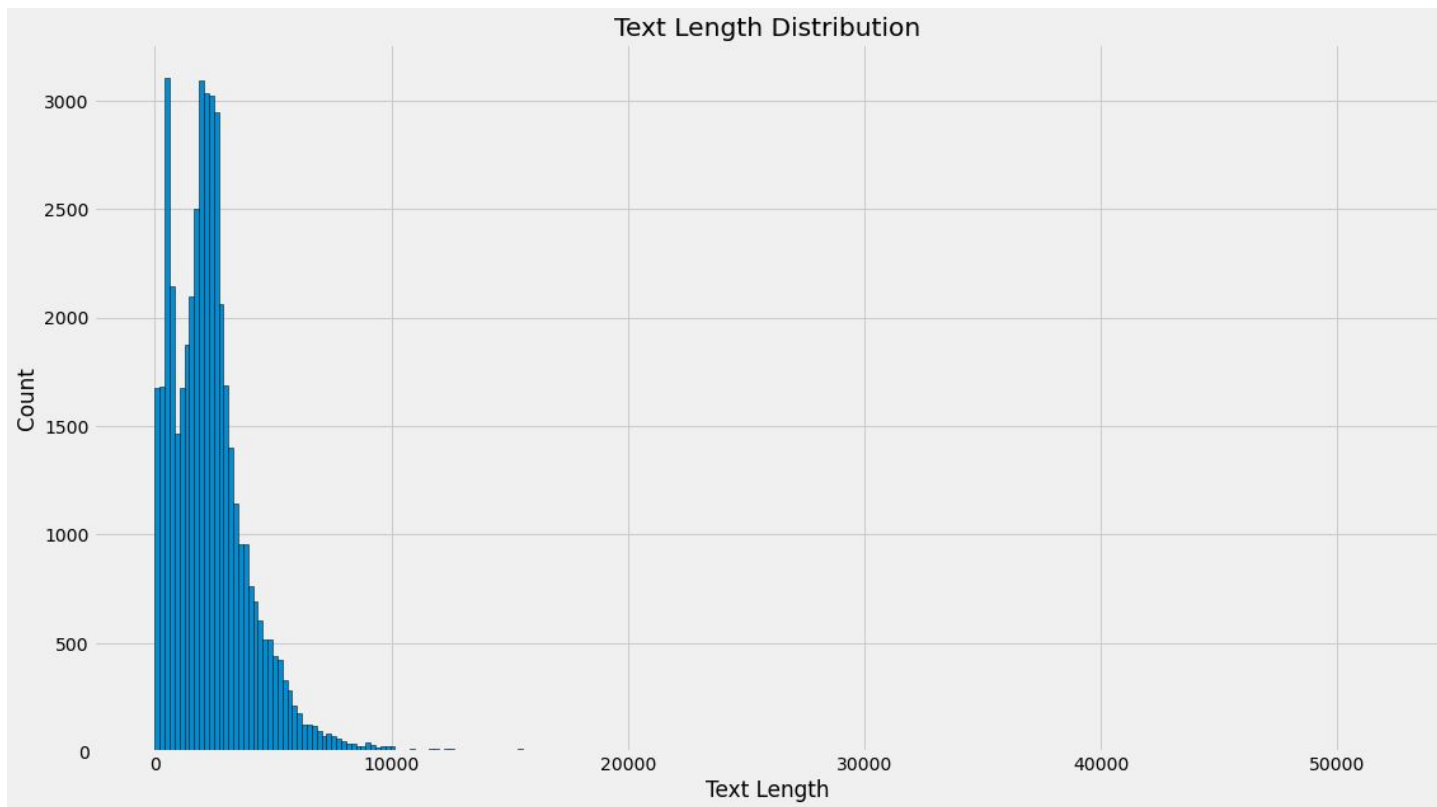
3. Feature Extraction

- a. CountVectorizer
- b. TF-IDF Model

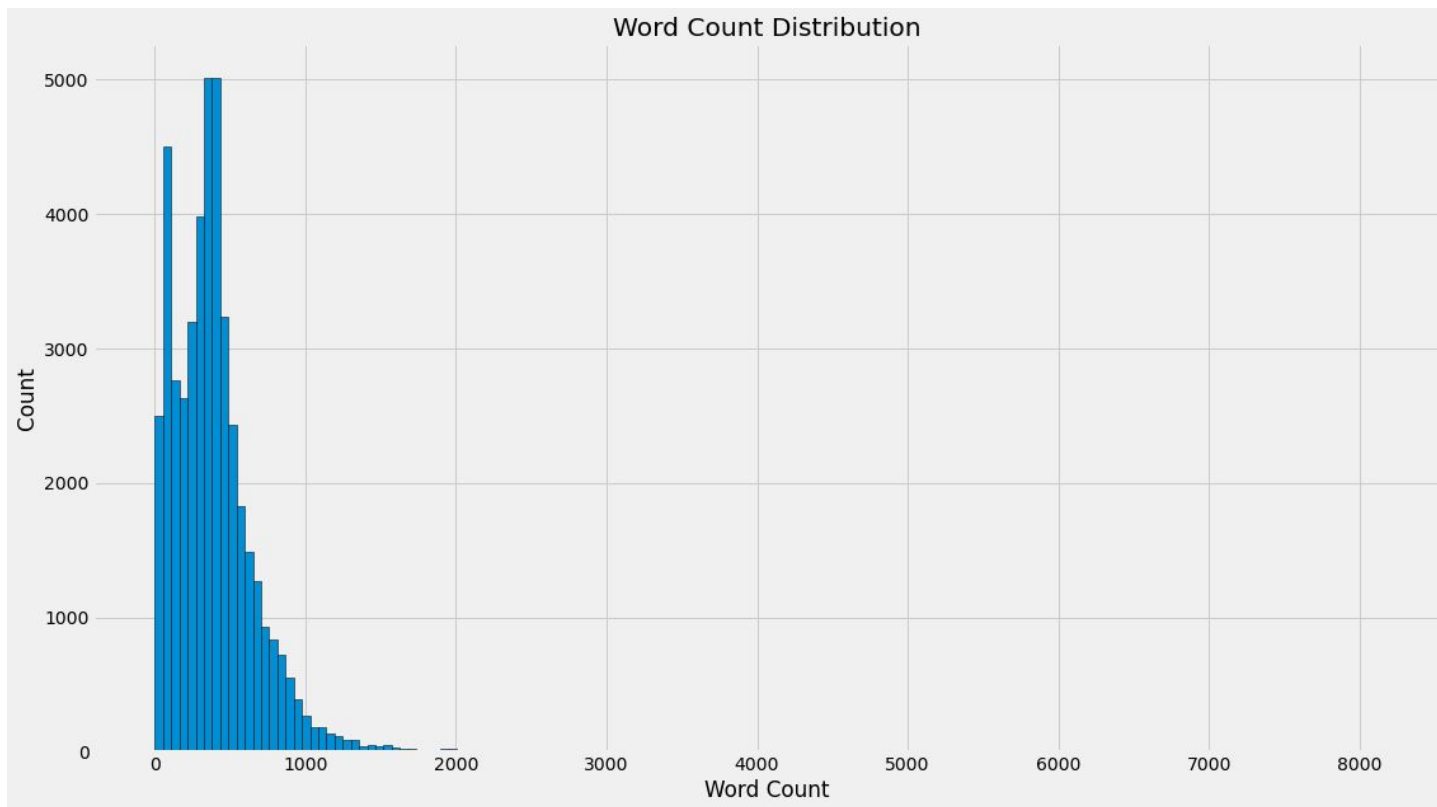
4. Split the Dataset

- c. Training Data (80%)
- d. Test Data (20%)

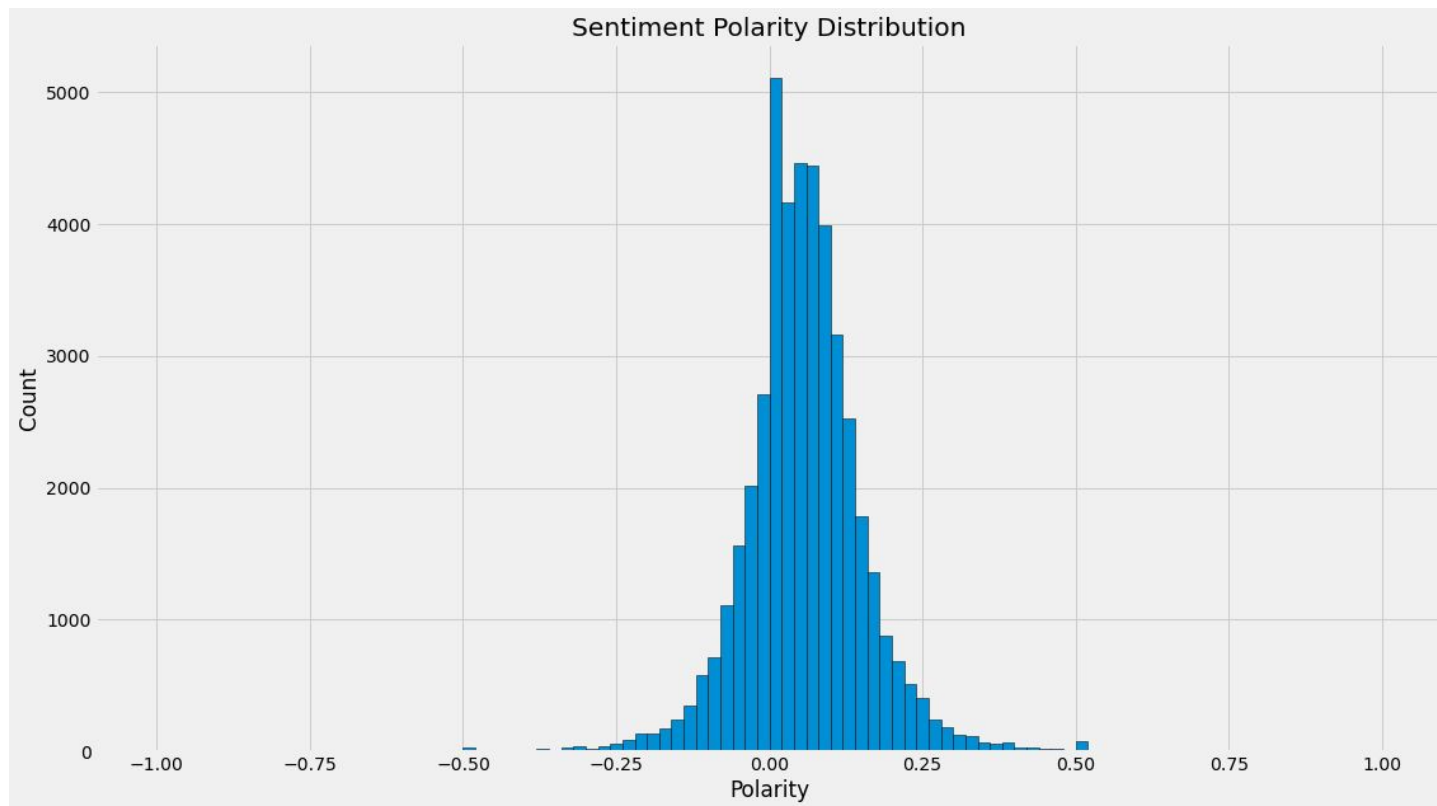
Text Length Distribution



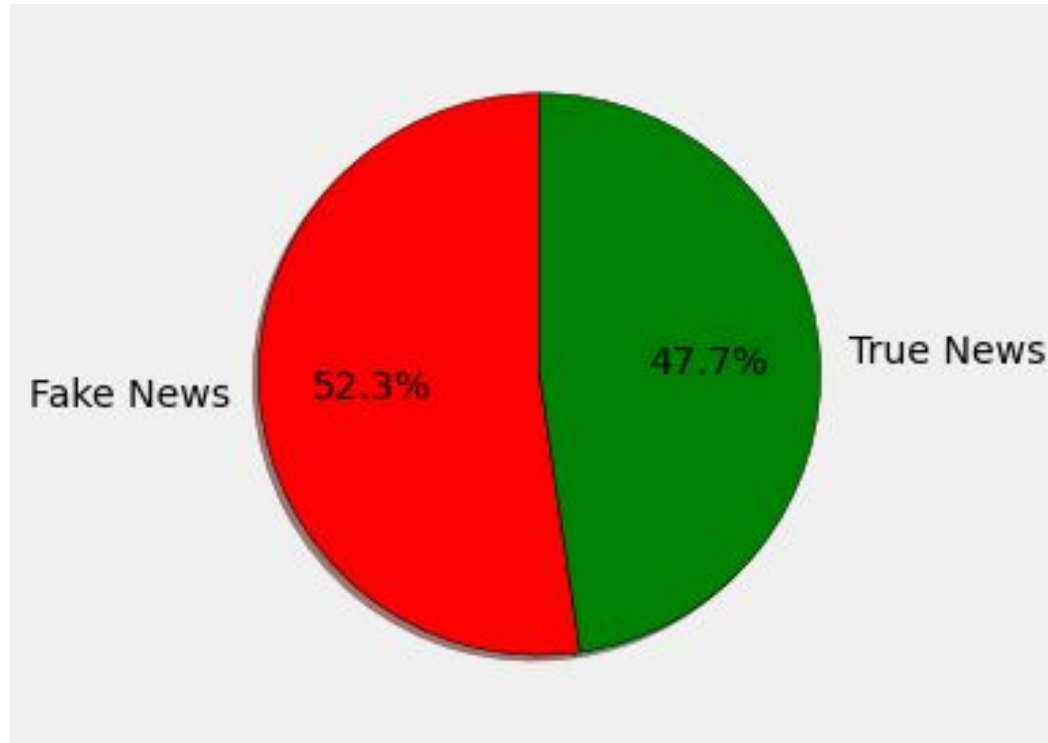
WordCount Distribution



Sentiment Polarity Distribution



Distribution of Binary Class

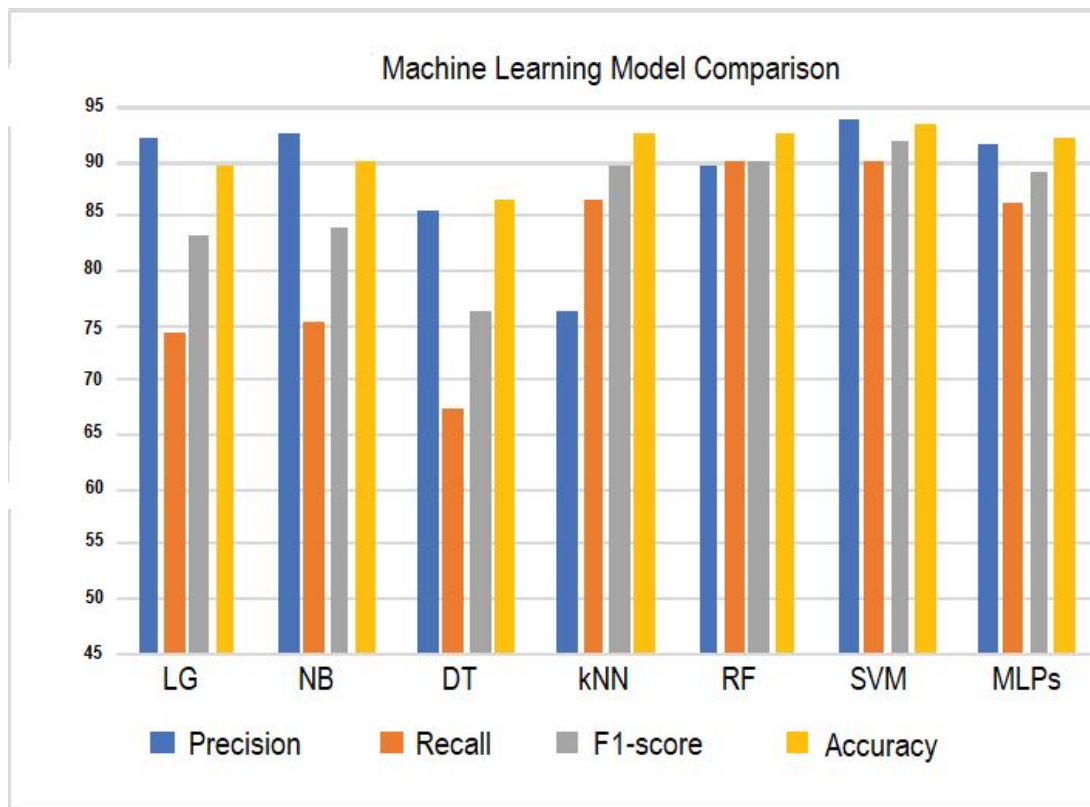


Machine Learning Implementation

Supervised Classification Techniques

1. Logistic Regression
2. Naive Bayes
3. Decision Tree
4. Random Forest
5. k-Nearest Neighbors
6. Support Vector Machine
7. Multilayer Perceptron

Result Evaluation



Application Design



The banner features a man in a grey shirt and blue pants leaning over a laptop. The laptop screen displays the word "NEWS" with a large red "FAKE" stamp over it. The background is a light beige color with decorative elements including various sized circles in shades of green, orange, and grey, some with dotted patterns. The text "BIG DATA ANALYTICS" is in large blue and red letters, followed by "PROJECT: FAKE NEWS DETECTION" in green and black, and "BY: AMIR ,JACEK, JB, JAVIER" in blue and black.

BIG DATA ANALYTICS
PROJECT: FAKE NEWS DETECTION
BY: AMIR ,JACEK, JB, JAVIER

Enter Your Comment Here

Predict

Further work

- Ensuring daily updating of the model
- Improving front-end interface:
 - Queries on individual statements
 - Queries on statistics
- Testing and ensuring maintainability

Questions
