By:
Rahaf Alqahtani
Dimah Albunayyih

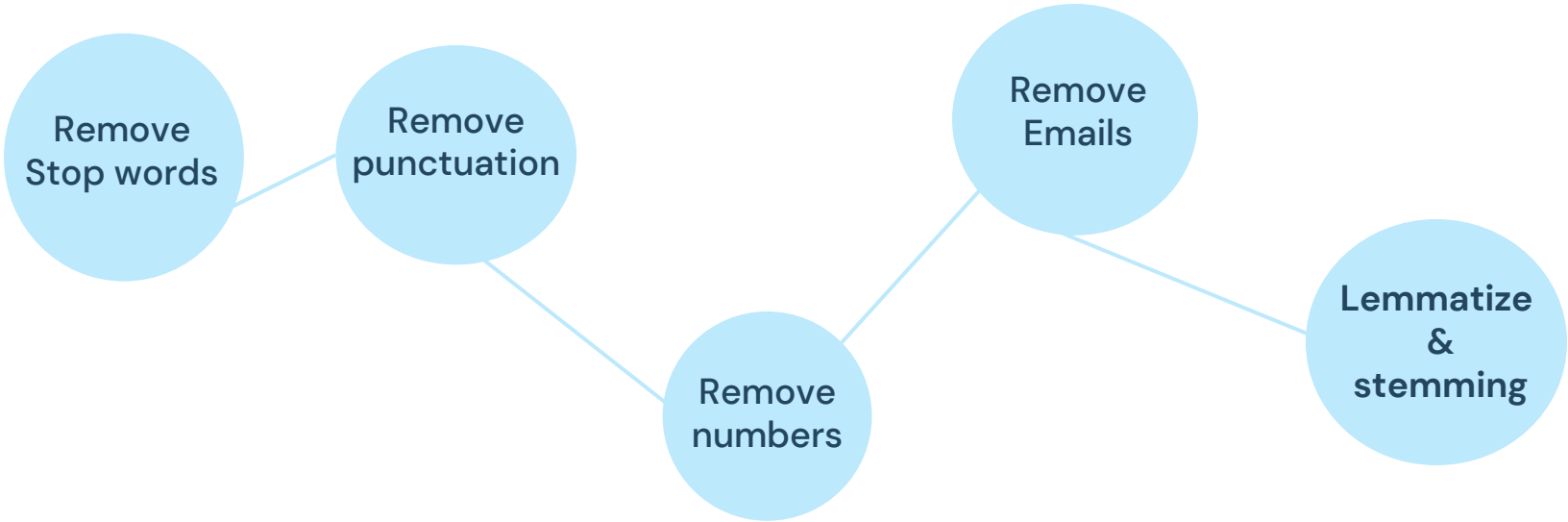
Table of Contents

Introduction

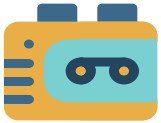
We all see a lot of fake news almost on a daily . Often its purpose is to spread fear and corruption, especially in societies where the level of digital awareness is low.

Workflow

NLP


Remove Stop words
Remove punctuation
Remove numbers
Remove Emails
Lemmatize & stemming

Dataset



News dataset

44k rows

4 columns

#13 01.01.2025

DAILY NEWS

WORLD NEWS TODAY

Economic News

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo egestas. Praesent vitae faucibus nisl. Interdum et malesuada fames ac ante ipsum primis in faucibus. In aliquam mauris sed risus fermentum porta. Curabitur quis dolor quis risus accumsan dictum. Nullam efficitur tellus eu est congue maximus. Donec vitae urna sodales enim venenatis vestibulum. Donec gravida varius lectus id accumsan. In tortor orci, molestie quis lorem ut, euismod molestie quam. Sed nec mauris nec urna efficitur pretium. Suspendisse quis viverra dui. Donec suscipit sem eu sem ornare, in rutrum orci condimentum. Curabitur scelerisque magna nisl, quis blandit enim dictum a. Nunc sollicitudin sollicitudin purus.

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo egestas. Praesent vitae faucibus nisl. Nam in urna sed orci gravida dapibus.

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor.

#13 05.01.2025

ECONOMIC NEWS

QUISQUE ENIM

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor.

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor finibus porttitor. Cras a lorem nec.

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo egestas. Praesent vitae faucibus nisl. Nam in urna sed orci gravida dapibus. Quisque posuere feugiat.

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor. Cras a lorem nec lectus imperdiet laoreet ut vitae nisl. Aliquam accumsan erat vel commodo egestas.

PROIN ELEIFEND FINIBUS MAURIS

Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor. Proin eleifend finibus porttitor.

Cras a lorem nec lectus imperdiet laoreet ut vitae nisl.

Nam Urna Sed Orci Gravida Dapibus

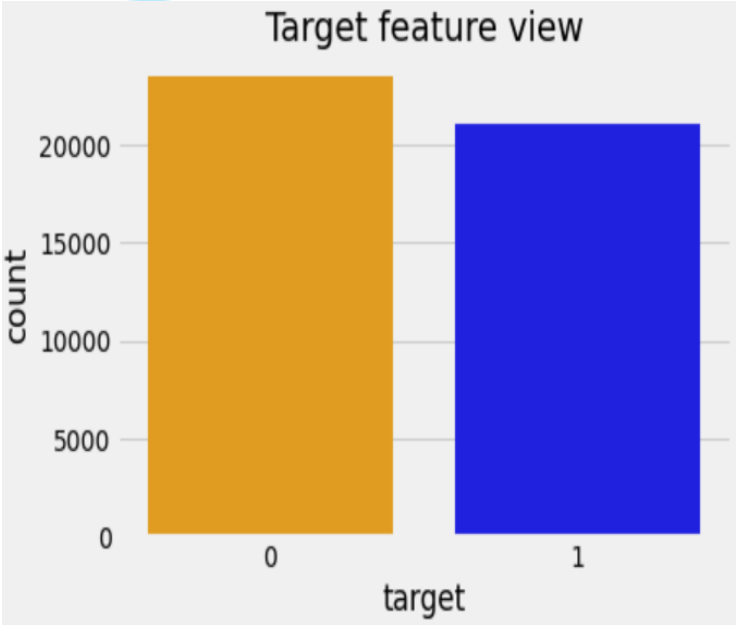
Lorem ipsum dolor sit amet, adipiscing elit. Phasellus cursus et enim ac scelerisque. Nam in urna sed orci gravida dapibus. Sed sed diam sollicitudin, bibendum dui et, facilisis mauris. Quisque posuere feugiat ullamcorper. Sed ornare nibh eu enim posuere, gravida cursus augue efficitur. Curabitur ut libero nunc. Cras mollis mi dolor.

Tools

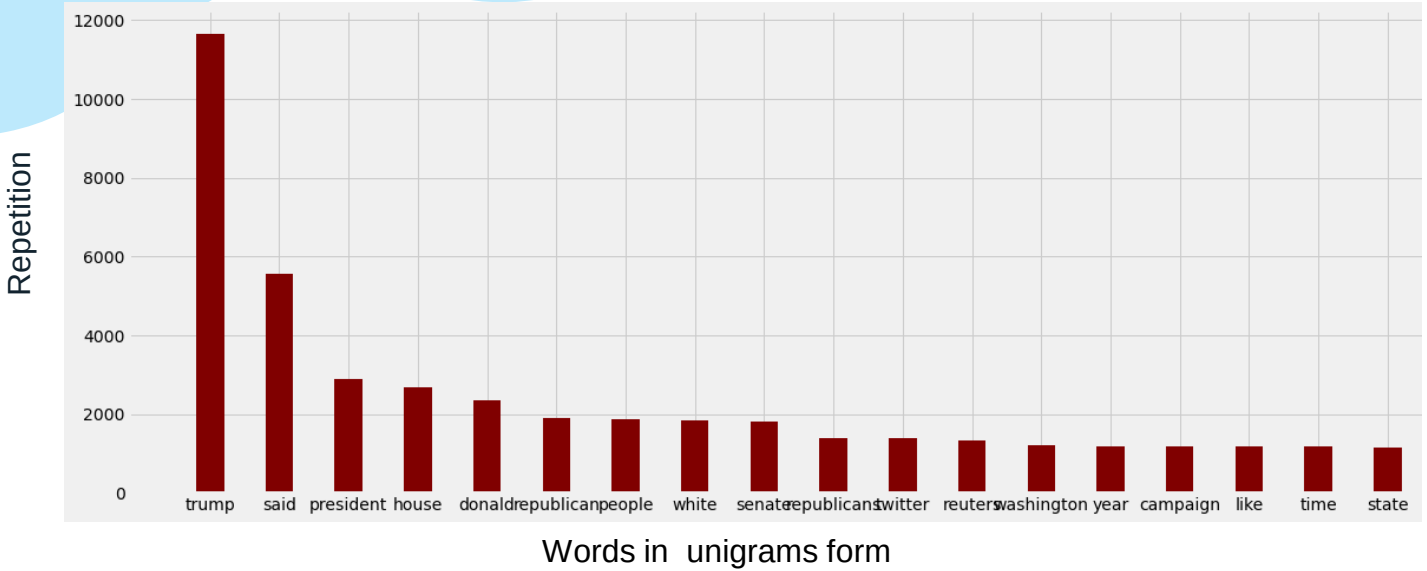
- Python, Jupyter notebook
- NumPy, Pandas
- Matplotlib, Seaborn and wordcloud
- Sklearn
- NLTK and gensim
- Regular Expressions
- Ngrams

EDA

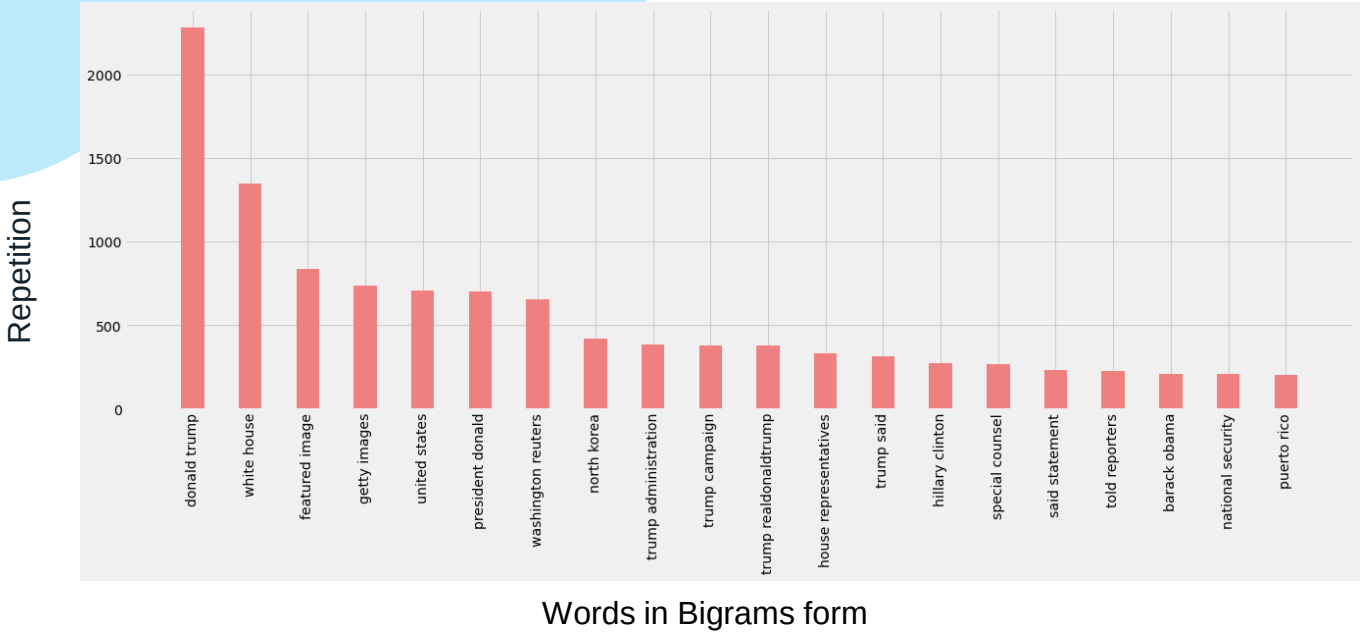
Fake and real counts



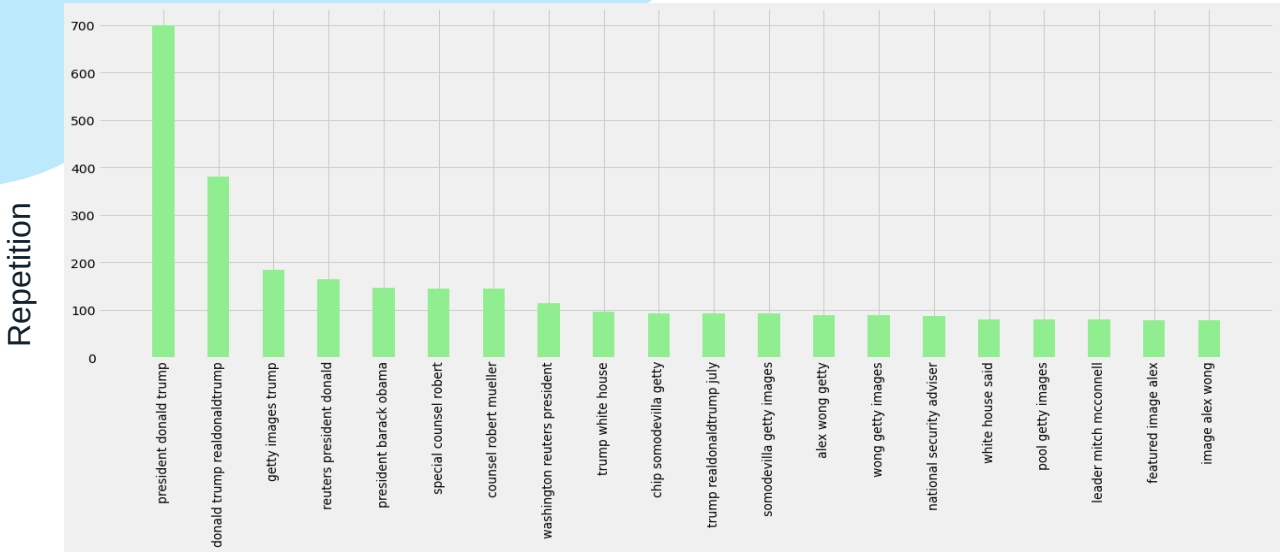
Unigrams Counts



Bigrams counts

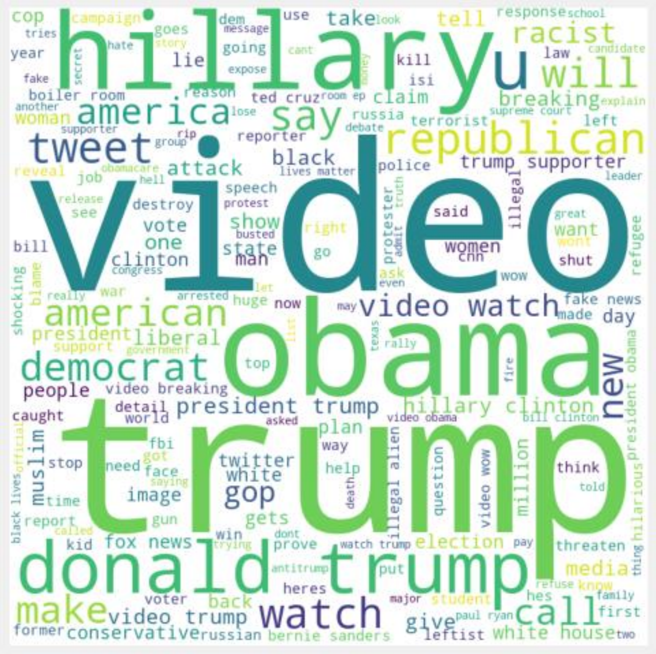


Trigrams Counts



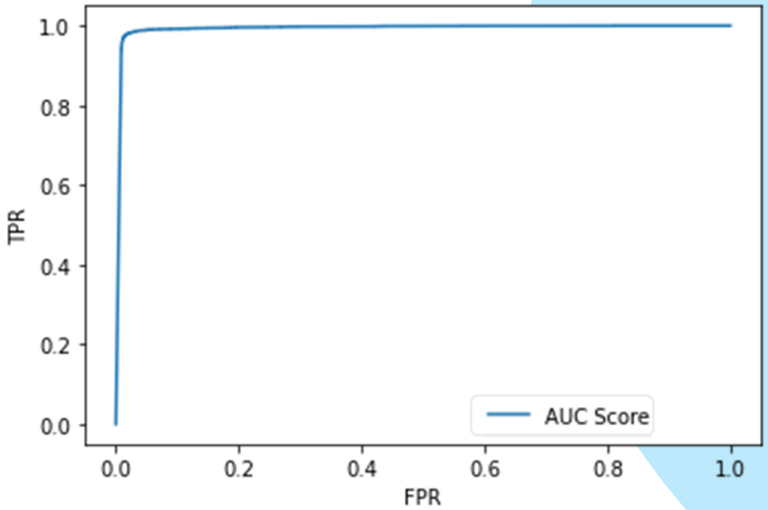
Most common words in real news

Most common words in fake news



Classification

Classification using naïve bayes models

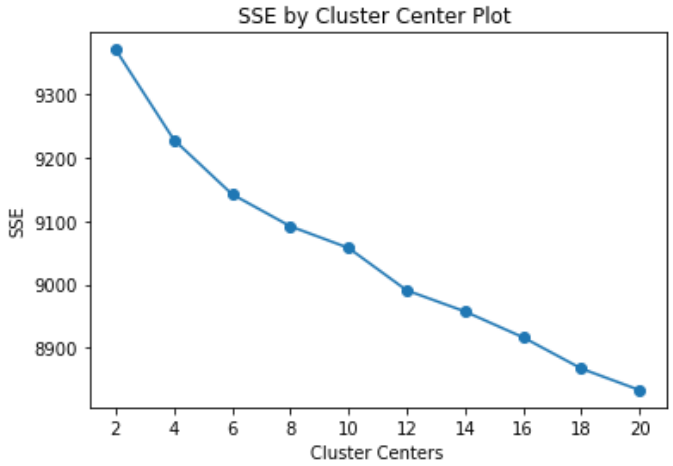


Accuracy Score	
0.95	**MultinomialNB** Baseline model
0.965	**BernoulliNB**
0.98	**BernoulliNB** After Tuning
0.978	**BernoulliNB** Testing

Topic Modeling

What is the optimal number of topics ?

K-means



LDA

Hyperparameter Tuning

10 Topics

LDA using different libraries

Bag of word

TF-IDF model

LDA

Count Vectorizer

LDA

Conclusion

- In this project we built a model that predict fake and real news by using classification technic
- The best score was 98%
- Our best method in topic modeling was LDA

Thank You