

# Natural Language Processing Fake and Real News

Rana Alzahrani | Shahad Alomayrah

## Abstract :

The goal of this project is to use Natural Language Processing and Unsupervised Learning models to categorize them into sections. After refining the approach, we will analyze the articles and categorize them into groups such as political.

## Design :

In this project, we are working on a dataset that consists of text about the News . Our observation is a Fake or Real News Building NLP model which is unsupervised learning that focuses on finding meaningful topics on News.

## Data :

data provided by kaggle:

- Max words number is 143034 words
- Min words number is 916 words
- Number of 20758 rows
- Number of columns 5
  - id: unique id for a news article
  - title: the title of a news article
  - author: author of the news article
  - text: the text of the article; could be incomplete
  - label: a label that marks the article as potentially unreliable
    - 1: unreliable 0: reliable

## EDA :

- Remove NULLS and Duplicate values
- Merge Title and Introduction columns
- Remove Unnecessary columns

## Preprocessing :

- Remove Non-English Char
- Remove extra spaces
- Remove Numbers
- Remove one litter
- Remove Arabic Punctuations
- Remove Arabic Stop Words

## Algorithms :

- **Embedding**
  - CountVectorizer
  - TF-IDF
- **Topic Modelling**
  - LatentDirichletAllocation(LDA)
  - LatentSemanticAnalysis(LSA)
  - Non-NegativeMatrixFactorization(NMF)
  - CorEx
  - Clustring

**Best model is NMF with Count Vectorizer** because it generates manful topics, but other models give unrelated terms for each topic

### **POSITIVE NEWS:**

pay, job, tax, money, hard, bad, economy **Topic 1 'Taxes'**

money, bank, debt, reserve, loan, trillion **Topic 2 'Financial'**

military, force, nuclear, attack, Iran, weapons **Topic 3 'Wars'**

Arabs , jewish , expel, attack, Palestinians , Zionists **Topic 4'Israel'**

fbi, investigation, case, evidence, source **Topic 5'Laws'**

vote, win, media, candidate, presidential, democratic **Topic 6' Election'**

## **NEGATIVE NEWS:**

women, life, feel, right, mother, include **Topic1 'Women Rights'**  
vote, political, win, presidential, conservative, candidates **Topic2 'Election'**  
kill, security, fight, bomb, strike, troop, minister **Topic3 'Military Strategy'**  
officer, gun, arrest, drug, suspect, gang, protest **Topic 4 'Police'**  
tax, pay, money, bank, billion, increase, deal, firm **Topic 5 'Financial'**  
trade, missile, deal, economic, global, weapons **Topic 6 'Trade'**

## **Tools :**

- Software Platform:
  - Jupyter Notebook
- Programming Language:
  - Python
- Python Libraries:
  - Sklearn
  - Nltk
- Data manipulation libraries
  - Pandas
  - Numpy
- Visualization
  - Matplotlib
  - Seaborn

## **Future Work :**

- Recommendation systems
- Remove More Stop Word