Fake and Real News

By:

- Rahaf Alqahtani
- Dimah Albunayyih

Problem

Description With the spread of social media and the ease of spreading news, there is a lot of fake news, which may cause many problems. So by using Natural language processing (NLP), classification, clustering and different techniques to provide an information that help to know whether the news is true or not.

Data Description

The data that will be used in this project has been taken from kaggle website (https://www.kaggle.com). It includes data such as: (Text, Subject, Date, etc.) for each news. It includes more than 44K columns and 4 features.

Algorithms

After we loaded our data, we transformed it into data frame, we have done the cleaning and text preprocessor, and then we did EAD .After that we used libraries in TOOLS to do Classification naïve base models then we did LDA using two libraries (gensim, sklearn) with tfidf and count vectorizer.

TOOLS

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

Communication

We will share our work in our GitHub accounts

- https://github.com/DimahAlbunayyih
- rahaftech/NLP_Unsupervised_Project (github.com)