Saudi Newspapers Articles

Abstract

The media occupies an important role in a democratic society, the press is not just a source of news, but a reference for current information and a tool for public criticism, it greatly influences the making of public opinion.And the press has many basic functions, and the nature of each newspaper differs from the other in terms of the personality of the writer.

This project will explore different methods of using NLP on the data with the goal of finding a good way of organising papers past by topic. We will visualise and report on the results and preprocess over time and by region to give a descriptive overview of P papers past through the topics. We will also make recommendations about the best ways to present topic model results to users.

Data

This data set from Github contains a collection of 3000 rows , 3 columns descriptive newspaper articles extracted from various Saudi newspapers on the Internet
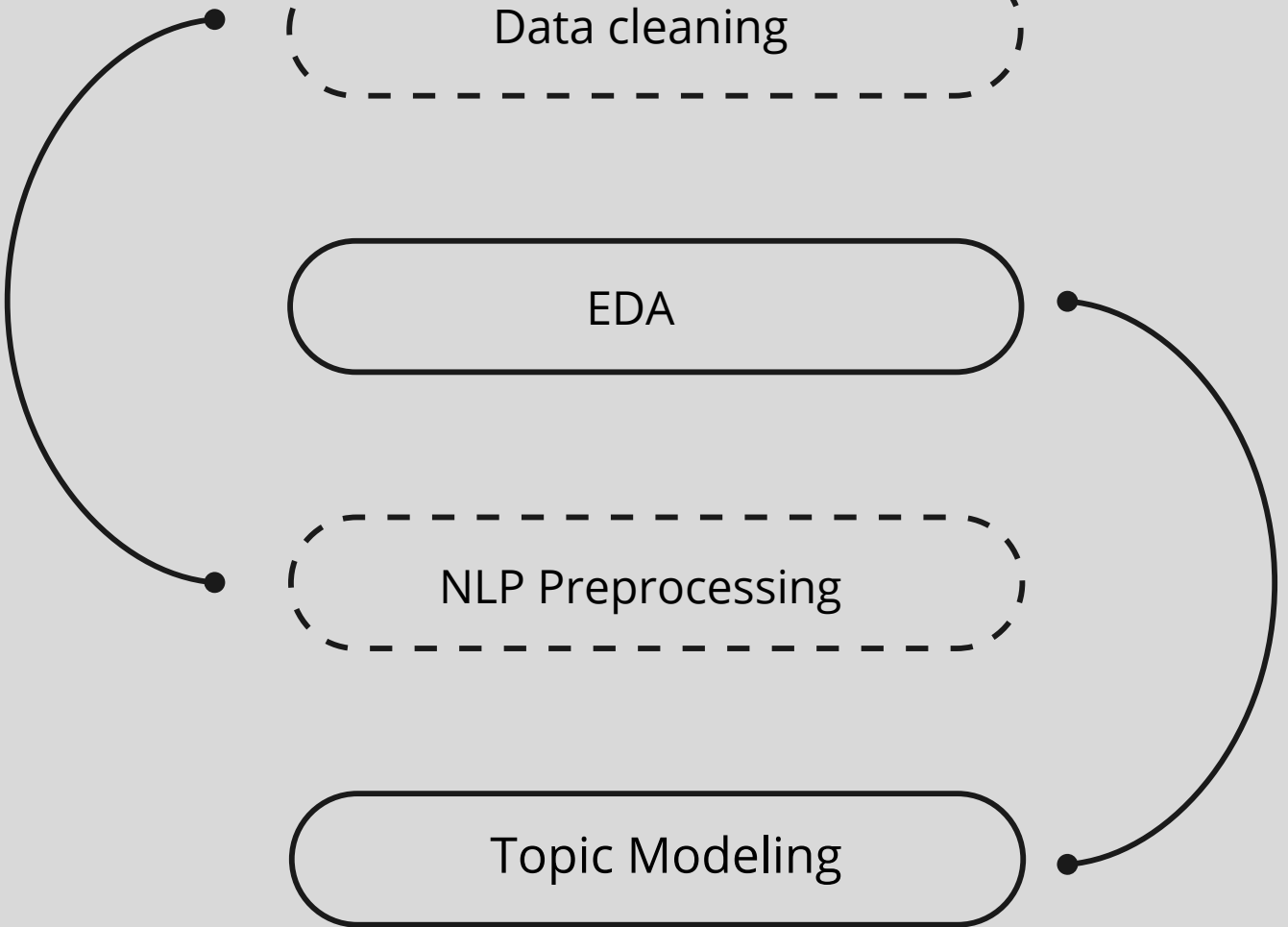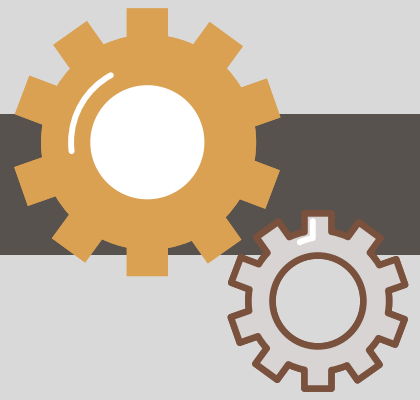
Algorithms

Understand the DataSet

Data cleaning

EDA

NLP Preprocessing

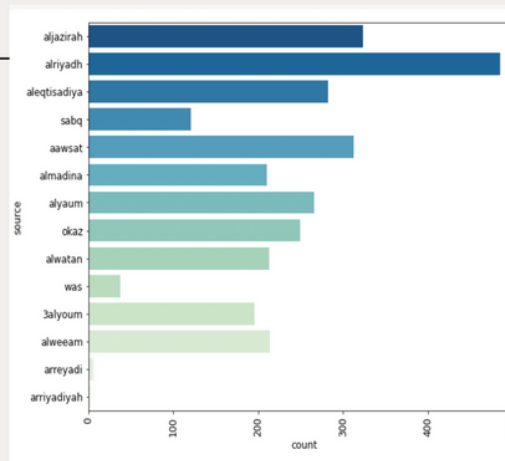Topic Modeling

Tools

Technologies

- Python
- Jupyter notebook
- Google colab

Libraries

- Pandas
- NumPy
- Matplotlib
- Seaborn
- Sklearn
- Nltk
- Arabic Libraries

Communication

This figure shows us what is the most widely reported in Saudi news.

Word cloud of the most frequent words

NMF (**Non-Negative Matrix Factorization**)

After reviewing 8 of the topics, it became clear to us that the best topic is 5 and contains different documents

The best choice

Public politics
Saudi politics
Sports

Public politics
Saudi politics
Sports
Economy

Public politics
Saudi politics
Sports
Economy
Entertainment