Analysing the trend and predicting the future of blockchain technology using twitter data

Aim: Analysis of blockchain data on twitter using Tweepy (Python library).

The aim of this project is to examine the blockchain discussions on twitter and analyzing the trend of blockchain and demonstrating the future of blockchain technology in multiple industrial applications. In order to identify and further the development of the blockchain technology, this project reviews the Twitter data on the blockchain using various Natural Language Processing (NLP) tools to extract knowledge from unstructured text data.

- 1. Extract real-time data from twitter
- 2. Pre-processing the data (lemmatization)
- 3. Clustering, Labelling, Topic modelling (LDA or NMF)
- 4. Cluster Map visualization
- 5. Visualization of trend of different domains

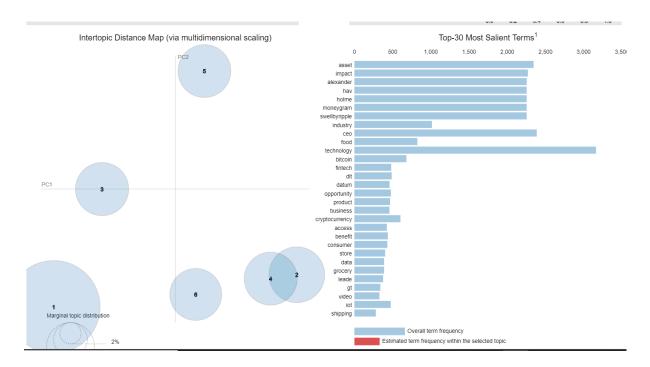


Fig 1. Overall term frequency and estimated term frequency within the selected topic

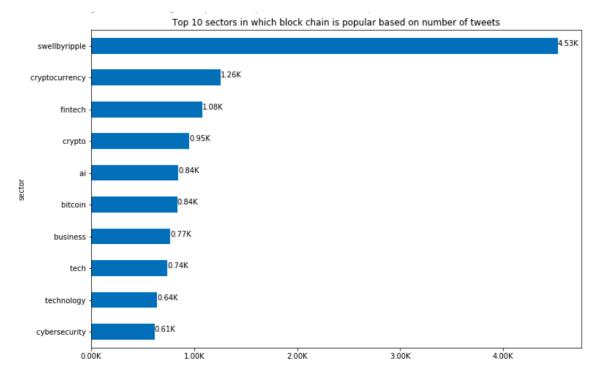
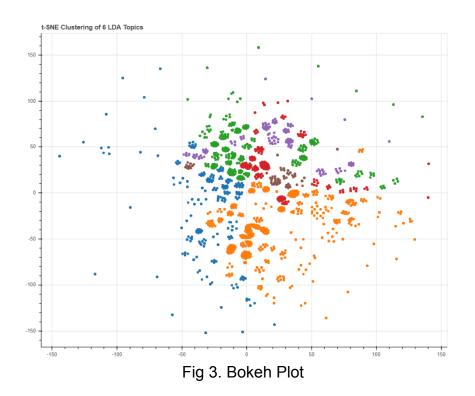


Fig 2. Number of tweets per sector



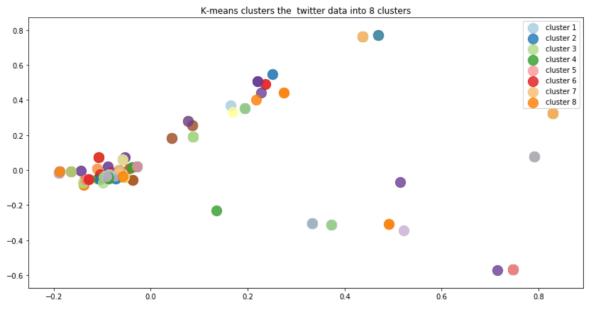


Fig 4. K means Clustering taking 8 clusters as an example

Topic 0 benefit dlt industry opportunity consumer_{tech} business product smartcontract leade

Topic 2 enhance risk finserv cryptocurrency efficiency gtaid_device iotfintech cecil_foundation

topic 4 europe data food access code grocery news datum Topic 1 amp trade adoption bitcoin gov finance technology video team

Topic 3 asset^{moneygram} hav technology impact^{healthcare} swellbyripple ceo alexander

Topic 5 gamechanger market startup technology cryptocurrency shipping solution iotblockchain week

Fig 5. Word cloud topic modelling example