Analysis of mails from Hillary Clintons Mail dataset

Debojyoti Dey (15511264), Nimisha Agarwal (15511267)

November 9, 2016

Overview

Problem Statement

- Extracting important topics from the e-mails.
- Getting an idea of US Foreign policy and relation with different nations.
- Persons linked to Clinton and Hillary's sentiment about them.

Sentiment Analysis

- Process of identifying and categorizing opinion.
- For determining the person's attitude towards a particular thing.
- Various libraries are available in python for this purpose.
- We have used VADER library in NLTK for this purpose.

Topic Modelling

- Technique for discovering the abstract "topics" in a set of documents.
- If a document is on a certain topic, then it is expected to contain particular words.
- We have used LDA library of gensim for this purpose.

Acievements

- Extracting important topics from the e-mails.
- Getting an idea of US Foreign policy and relation with different nations.

Couldn't Achieve

- Persons linked to Clinton and Hillary's sentiment about them.

Demo

Conclusion

References

📄 https://www.kaggle.com/kaggle/hillary-clinton-emails