Project Proposal

Group Name: Neem
Group member: Yuanzhi Xu, Chia-wei Lin, Wen Xie
Project Name: Sentiment analysis on COVID-related tweets

Project description

This project carries out a sentiment analysis on people's emotional attitude (positive, negative and neutral) towards Covid-19 by gathering and analyzing real world Tweets text, aiming at generating insights on public reaction towards the pandemic within the time period from 01/2022 to 06/2022. Natural language processing methods and data mining techniques are to be applied.

The public's attitude towards Covid-19 will be a factor of great weight for the government and organizations to lay or lift certain regulations towards the pandemic.

Plan to approach the project

- Dataset:
    a. Sentiment140 dataset with 1.6 million tweets
    https://www.kaggle.com/datasets/kazanova/sentiment140
    This dataset contains 1,600,000 tweets extracted using the twitter api . The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment. This would be used as the train dataset since it is clean and already has target value.
    Using the model developed from this dataset, we can do sentiment analysis and prediction on real world data collected from Tweets text:
    b. Collect real-world tweets
    https://github.com/echen102/COVID-19-TweetIDs

- Libraries: Pandas, Numpy, Matplotlib, Scikit-learn, Pytorch, GloVe, Textblob, NLTK, WordCloud…
- Methodology: the project will be carried out in the following steps:

1. Data collection
2. Data cleaning: Text preprocessing
3. Data analysis
4. Model Training
    a. BernoulliNB Model
    b. LogisticRegression Model
    c. Decision Tree Model
    d. SVM Model
    e. LSTM Model
5. Model Evaluation: Accuracy and Confusion Matrix
6. Collect real-world tweets
    a. https://github.com/echen102/COVID-19-TweetIDs
    b. <2022/01 - 2022/06>

7. Classify real-world tweets
8. Interpret the result

Final deliverable

A project report that analyzes and predicts the public's sentiment towards Covid-19 from 01/2022 to 06/2022.