Project: Twitter Sentimental Analysis System

Members: Vaibhav Pandey (SID: 18103072)

Nikhil Jaglan (SID: 18103118)

Level 0:

*Process:* Twitter Sentiment analysis System.

*Description:* This using the access token secret from twitter API acquires the tweets made by users. Thereafter processing and classification of tweets is carried out. The trends thus obtained are send/outputted to the analyst(system user) for any future analysis.

*Input : Tweets*

*Access token*

*Access token secret*

*Output : Sentiment trends*

*Logic Summary: We first request the twitter to grant us access to tweets of people .On approval it grants us Access token secret .We then use it to acquire the tweets. We give these to analyst who performs data cleaning and data processing and using various ml models tries to predict the best model for getting most accurate results. We then display/visualize our results.*

*Entities: User of twitter*

*Data Flow Name: Twitter API Connectoin*

*Description*: We request twitter to connect to twitter API

*To Process:* Twitter API

*From Process: Sentiment analysis system*

*Data structure:* access token and access token secret

*Data Flow Name:* Getting tweets

*Description*: We try to get the relevant tweets related to politics.

*To Process:* Users

*From Process: Analyst/Sentiment analysis system*

*Data structure:* Tweets

*Data Flow Name:*  Sentient trends to analyst

*Description*: Doing data processing and visualisation

*To Process:* *Analyst*

*From Process: Sentiment analysis system*

*Data structure:* Processing result

Level 1.1:

*Process1: Data Collection System*

*Description1:* This acquires the tweets of people from the twitter database with the help of access token and access token secret given.

Input: Keywords for Tweets extraction

Access token and Access token secret

Output: Tweets Data

Local database

Logic Summary1: We request the twitter for unique access token secret . After it's acquisition we extract the tweet's made by people related to politics with the help of proper keywords. We update our local database with the extracted tweet's and sent the tweet's data for data processing to Level 2.

*Datastore*: Local database is used

Item fields:

1. Access token : alphanumeric string

2. Access token secret : alphanumeric string

1.Tweets : Type String . Contains the tweets/text tweeted by people

*Data Flow Name: Twitter API Connection*

*Description: Connecting with the twitter API*

*To Process:* Twitter API

*Data structure:* *Access token and access token secret*

*Data Flow Name:* Tweets Extraction

*Description*: Acquiring the tweets

*To Process:* Users

*From Process: Twitter API connection*

*Data structure: Tweets*

*Data Flow Name: database filling*

*Description*: *Tweets movement to database*

*To Process:* Data Processing

*From Process: Tweets Extraction*

*Data structure:* tweets

Level 1.2:

*Process2: Data Processing System*

*Description2:This is related to the processing of data and finding the optimal hyper-parameters*. We finally train our model on the learnt parameters.

Input2: Tweets Data

Local Database

Output2:

Data processing result

*Logic summary2:Here we first try to remove the various stop-words and irrelevant information. We then try to extract most relevant words and words which have maximum frequency and also define various thresholds for this and try to clean up the data. We then create a convolution layer model* and try to find optimal number of hyper-parameters by using validation

*Datastore*: Result database is used

Item fields:

1.Result : Type tuple . Contains the sentiment, it's degree and id (positive/negative/neutral and it's degree)

*Data Flow Name: Tweets from database*

*Description:* Getting the tweets from database i.e new and old all tweets

*To Process:* Data Processing

*From Process: Local Database*

*Data structure:* tweets

*Data Flow Name: Tweets from Data Collection*

*Description:* Getting the tweets from previous level

*To Process:* Data Processing

*From Process: Data Collection*

*Data structure: tweets*

*Data Flow Name: Tweets from database*

*Description:* Getting the tweets from database i.e new and old all tweets

*To Process:* Data Processing

*From Process: Local Database*

*Data structure: tweets*

*Data Flow Name: Result storage*

*Description:* Storing the results of sentiment analysis

*From Process: Data Processing*

*Data structure:* Processing result

Level 1.3:

*Process3: Data visualisation*

*Description3:Here we try to visualise our results, accuracy and the loss function*

Input3:Data processing result

Output3: Sentiment trends visualisation

Logic Summary : Here we visualise the sentiment trends obtained for proper analysis of trends obtained.

*Datastore*: Visualisation database is used

Item fields:

1.Result : Type graph . Contains the data visualisation graphs

*Data Flow Name: Visualisation result*

*Description: Gives the visualised results in database*

*To Process: Data processing*

*From Process: Data Visualisation*

*Data structure : Processing result*

-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*-\*