SENTIMENT ANALYSIS

ORGANIZATION – INDIAN SPACE AND RESEARCH ORGANISATION (ISRO)

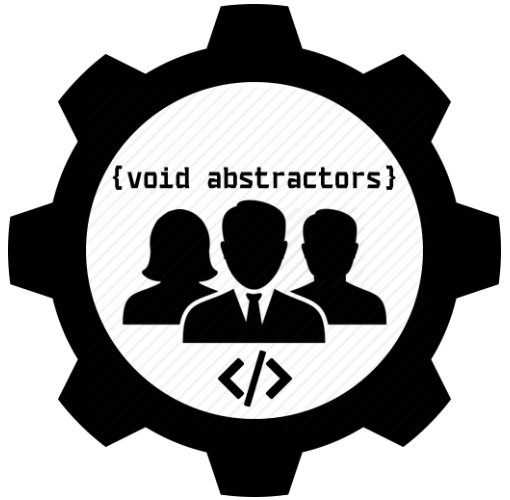
CATEGORY – SOFTWARE

DOMAIN BUCKET – MISCELLANEOUS

PROBLEM CODE – NM396

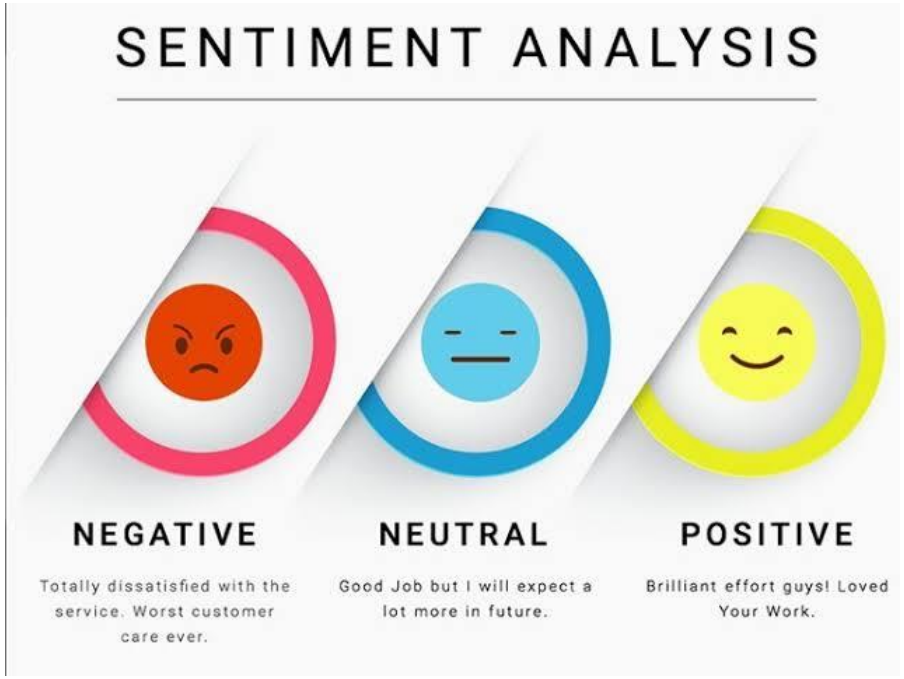
THE TEAM – VOID ABSTRACTORS

- TEAM LEADER – G Mukkesh [17BCE1128]
- TEAM MEMBER – Rohit Subramanian [17BCE1291]
- TEAM MEMBER – Amrit Krishna O [17BCE1133]
- TEAM MEMBER – Sanjana Dulam [17BCE1068]
- TEAM MEMBER – Akshay Kumar [17BCE1290]
- TEAM MEMBER – Maheshvar C [17BCE1172]



PROBLEM STATEMENT

- Sentiment Analysis using text feedback



THE GOAL

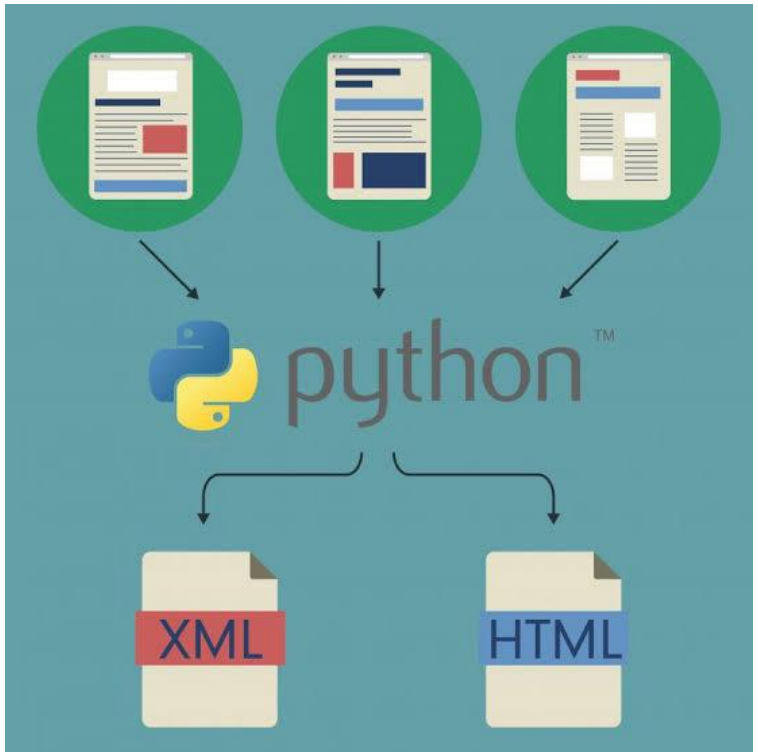
Sentiment Analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

The main goal is to go through all the feedbacks and categorize opinions expressed in the feedback forums. Essentially it is a software that can be used for feedback and review management system. This software must provide :

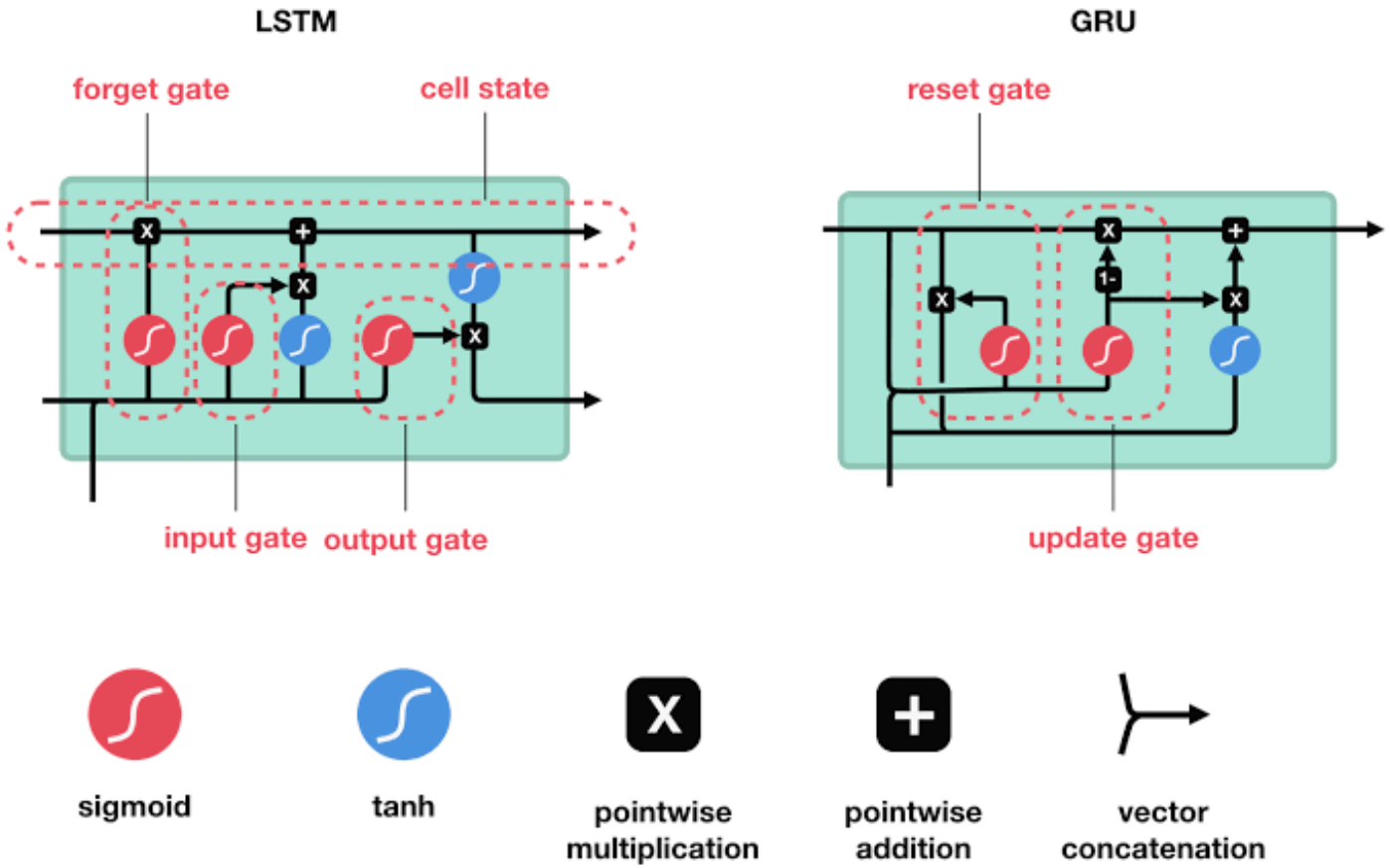
- Classification of individual comments/reviews.
- Determining overall rating based on individual comments/reviews.

TECHNOLOGY STACK

- **Web scraping** is the process of gathering information from the Internet. Even copy-pasting the lyrics of your favorite song is a form of web scraping! However, the words “web scraping” usually refer to a process that involves automation .
- The incredible amount of data on the Internet is a rich resource for any field of research or personal interest. To effectively harvest that data, you’ll need to become skilled at web scraping. The Python libraries requests and Beautiful Soup are powerful tools for the job.
- Beautiful Soup is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping. It is available for Python 2.7 and Python 3.



RNN's are good for processing sequence data for predictions but suffers from short-term memory. LSTM's and GRU's were created as a method to mitigate short-term memory using mechanisms called gates. Gates are just neural networks that regulate the flow of information flowing through the sequence chain. LSTM's and GRU's are used in state of the art deep learning applications like speech recognition, speech synthesis, natural language understanding, etc.



OUR PLAN

- Our plan is to use deep learning models to predict and categorize opinions or feedbacks regarding the website.
- We have trained a custom CNN – GRU – LSTM model with an accuracy of 92% on unseen test reviews
- A pipeline was set up to tokenize and embed the transformed text into a glove embedding model.
- The prediction is accelerated through the use of GPU which gives almost instantaneous results.
- Alternatively traditional ML models were also built which yielded an accuracy of 88% . (Lightweight Portable Models)

MODEL ARCHITECTURE

```
In [13]: model.summary()
Model: "model_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_6 (InputLayer)         (None, 800)               0
_________________________________________________________________
embedding_6 (Embedding)      (None, 800, 300)          27000000
_________________________________________________________________
conv1d_4 (Conv1D)            (None, 797, 16)           19216
_________________________________________________________________
dropout_4 (Dropout)          (None, 797, 16)           0
_________________________________________________________________
bidirectional_7 (Bidirection (None, 797, 40)           5920
_________________________________________________________________
bidirectional_8 (Bidirection (None, 797, 40)           7320
_________________________________________________________________
max_pooling1d_4 (MaxPooling1 (None, 398, 40)           0
_________________________________________________________________
flatten_4 (Flatten)          (None, 15920)             0
_________________________________________________________________
dense_6 (Dense)              (None, 30)                477630
_________________________________________________________________
dense_7 (Dense)              (None, 1)                 31
=================================================================
Total params: 27,510,117
Trainable params: 27,510,117
Non-trainable params: 0
_________________________________________________________________
```

FEASIBILITY STUDY

For the Feasibility Study , we would like to answer 3 basic questions.

- **Problem Statement Validity** : To go through all the feedbacks can be a tedious job. This software makes it easy to categorize opinions expressed in feedback forums and can be used for feedback management systems.
- **Social and Ethical Concerns** : Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation, and understand customer experiences.
- **Organization Capabilities** : Our team of 6 members are highly talented and capable and will be able to develop this software efficiently .

USE CASE / FUNCTIONALITIES

Some applications are:

- Social media monitoring
- Voice of customer (VoC)
- Customer service
- Market research

Custom trained GRU – LSTM model with an accuracy of 92%.

- Used for movie reviews / product reviews etc .
- Can be modified to show sentiment rating (on a scale of 0-10)
- By using transfer learning this model can be modified for cyber bullying detection / abusive language / insincere question classification.

LITERATURE REVIEW

The following research papers were reviewed:

- Sentiment analysis and opinion mining [Bing Liu]
- Malware classification with LSTM and GRU language models and a character-level CNN [Ben Athiwaratkun ; Jack W. Stokes]
- GloVe: Global Vectors for Word Representation [Jeffrey Pennington, Richard Socher, Christopher D. Manning]