Contents:

---

1. SAJAL [Kaggle_Profile]

Your Guide to Natural Language Processing (NLP)

The 7 NLP Techniques That Will Change How You Communicate in the Future (Part I)

A Practitioner's Guide to Natural Language Processing (Part I) — Processing & Understanding Text

Natural Language Processing is Fun! - Adam Geitgey

Word Embeddings for NLP

Beyond Word Embeddings Part 2

Intuitive Guide to Understanding GloVe Embeddings

https://github.com/dipanjanS/practical-machine-learning-with-python/tree/master/bonus%20content/nlp%20proven%20approach

https://github.com/joshzwiebel/Tweet-Sentiment-Extraction/blob/master/tweet_sentiment_extraction/Exploration/exploration.ipynb

New:
BERT Explained: State of the art language model for NLP
We are using a pre-trained model of this in kaggle. The paper was having 5000+ citations :)
Paper - https://arxiv.org/pdf/1810.04805.pdf

Kaggle notebook - https://www.kaggle.com/mohannksr/tensorflow-roberta-cnn-head-lb-v2

Introduction to sentiment analysis: What is sentiment analysis? |

---

2. SAURABH [Kaggle_Profile]

- Understanding the Kaggle competition: Tweet Sentiment Extraction
    - The maximum team size is 5.
    - Submission Limits
        - You may submit a maximum of 5 entries per day
        - You may select up to 2 final submissions for judging
    - Competition Timeline
        - Start Date: March 23, 2020
        - Team Merger Deadline: May 26, 2020
        - Entry Deadline: May 26, 2020
        - End Date (Final Submission Deadline): June 2, 2020 11:59 PM UTC

- **Notebook:**
    - Tweet Sentiment - Insight EDA
        - Basic understanding of EDA
        - Well written and explained
    - TensorFlow roBERTa + CNN head - LB v2
        - Can be used as backbone for our model
-
    -

---

3. ATUL [Kaggle_Profile

4. SAAD [Kaggle_Profile]

Reading done:

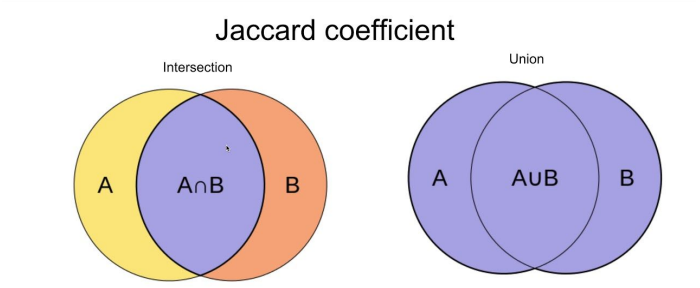
1.Understood everything about Kaggle notebooks: how to use,how to find datasets, how to upload datasets, how to download files from kaggle notebooks.

2.https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1 (Your Guide to NLP).Studied about Jaccard Score, Tokenization, Stemming, Lemmatization.

3.https://towardsdatascience.com/word-embeddings-for-nlp-5b72991e01d4. (Word Embeddings for NLP).Studied about word2vec.

---

5. Technical Definitions/ Summary of readings (Put bullets in your section above and explain briefly over here, please feel free to add extra points if you can find)

- **Jaccard Score**
  - 

  - $$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$
  - To check how much the predicted words matches with the actual words
  - Does not take care of meaning of the sentence, just the words in it.
- **Stop Words**
  - A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore.
  - Don’t want these words to take up space in our database
- **ngrams**
  - a set of co-occurring words within a given window (refer)

---

5. Meeting Update

- **May 21 (To be discussed)**
    - Discussion on Kaggle Notebook
        - Tweet Sentiment - Insight EDA
        - ..
    - Update your respective document and explain it
    - SNT evaluation (next weekend before May 30)
        - Prepare documentation well
            - Problem Statement
            - Plan of Action
            - Git repo (Kaggle to Git repo)
            - More focus on understanding the problem and main solution approach
    -