SEMI SUPERVISED LEARNING

PARTICIPANTS: DHEERAJ NAGURU
RIDHIMA CHEBOLU
ROHITH SHASHI

TABLE OF CONTENT

INTRODUCTION

- LARGE AMOUNT OF DATA
- SUPERVISED LEARNING : Labelling all the data
- UNSUPERVISED LEARNING : Use of Unlabeled Data
- SEMI SUPERVISED LEARNING : Supervised Learning + Unsupervised Learning

SEMI SUPERVISED LEARNING METHODS

- SELF TRAINING : Use of Pseudo Labelling Data.
- GENERATIVE MIXTURE MODELS : Use of only one labelled example per component.
- CO-TRAINING : Each classifier classifies the unlabelled data and teaches the other classifier with the few unlabelled examples.
- TRANSDUCTIVE SVM : Use of Hyperplane to maximize the margin.
- GRAPH BASED MODEL : Aims to classify unLABELLED data by learning the graph structure and LABELLED data jointly.

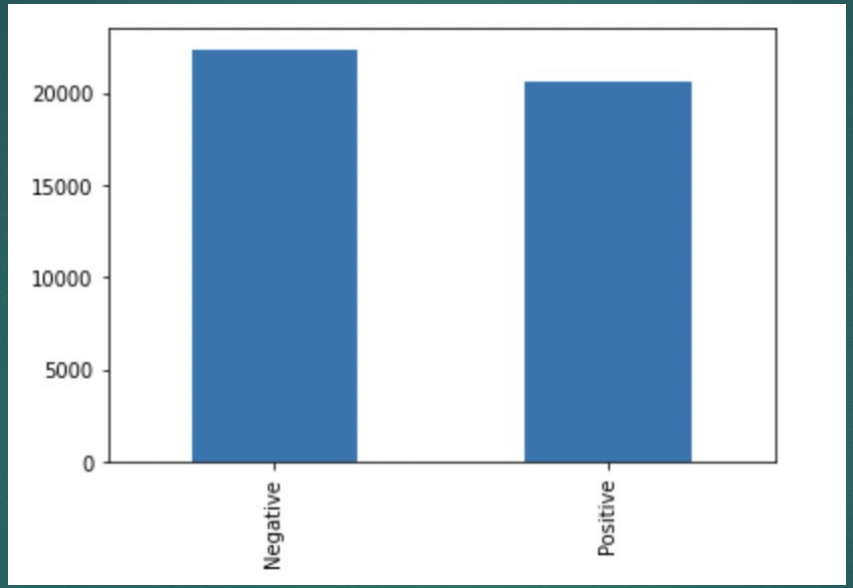
TWITTER SENTIMENT ANALYSIS

- PROVIDES INSIGHTS ON USER SENTIMENTS
- TWITTER HAS LARGE AMOUNT OF DATA
- NOT PRACTICAL TO LABEL ALL TWEETS
- USE OF UNLABELLED TWEETS FOR MODEL TRAINING
- SELF TRAINED SENTIMENT ANALYSIS

DATA PRE-PROCESSING

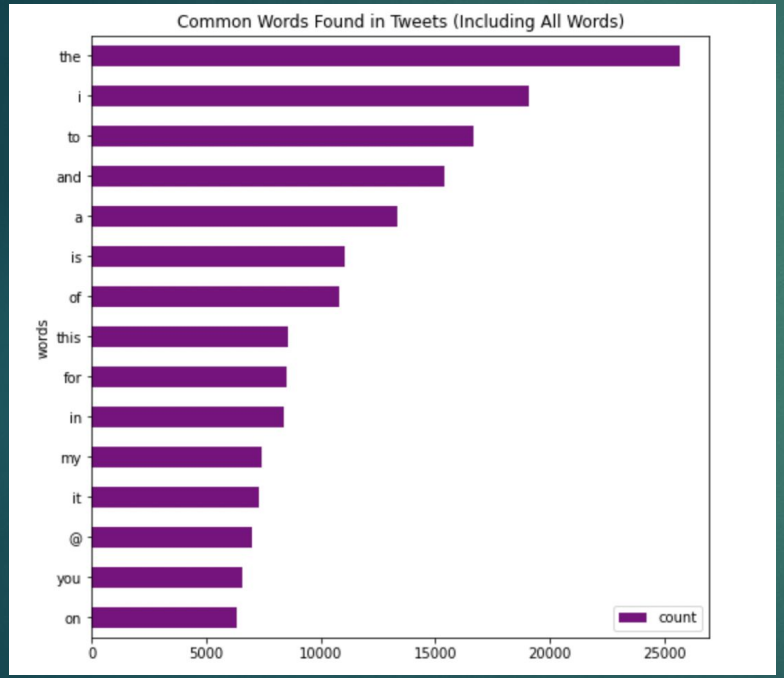
- DATA CLEANING
  - REMOVED URLs,
  - REMOVED '#' FROM HASHTAGS
  - REMOVED EMOJIS,
  - REMOVED HTML(OR SIMILAR) TAGS
  - REMOVED DUPLICATES
- REMOVED STOP WORDS
- TOKENIZATION
- STEMMING

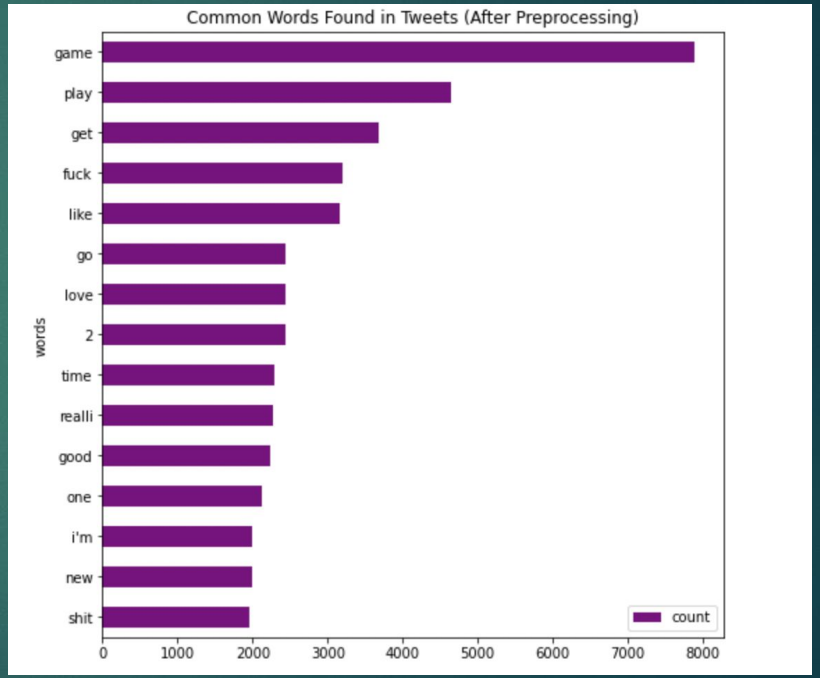
Overview of Dataset



Number of datapoints with positive and negative sentiments

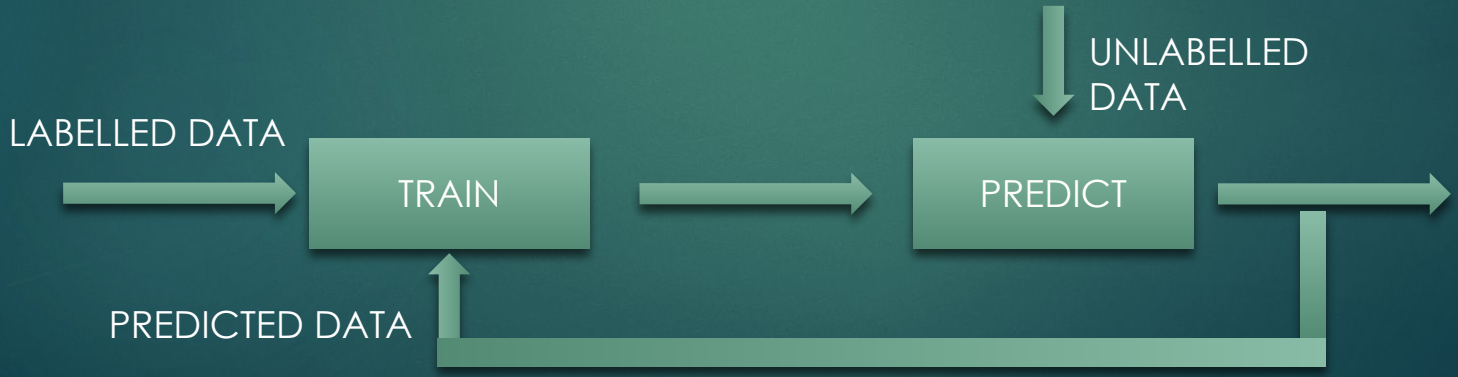
Overview of Dataset





SELF TRAINING

- CLASSIFIER TRAINED ON SMALL AMOUNT OF LABELLED DATA
- UNLABELLED DATA IS PREDICTED
- PREDICTED LABELS USED FOR TRAINING
- CLASSIFIER IS RE-TRAINED
- USES IT'S OWN PREDICTION TO TEACH ITSELF
- UNLEARN WHEN PREDICTION CONFIDENCE DROPS BELOW A THRESHOLD



COMPARISONS

- SUPERVISED LEARNING:
  - LOGISTIC REGRESSION
- SEMI SUPERVISED LEARNING:
  - SELF TRAINING
- UNSUPERVISED LEARNING:
  - K -MEANS

EVALUATION:

ACTUAL \ PREDICTED		
	TP	FN
	FP	TN

ACCURACY: $\frac{TP+TN}{TP+FP+TN+FN}$

PRECISION: $\frac{TP}{TP+FP}$

RECALL: $\frac{TP}{TP+FN}$

Upcoming Semester Focus Areas

- Co-Training method in semi supervised learning
- Longest Common Subsequence
- Documenting the Research

REFERENCES

- Xiaojin Zhu (2005), Semi-Supervised Learning Literature Survey
- V. Jothi Prakash , Dr. L.M. Nithya (2014), A Survey On Semi-Supervised Learning Techniques, International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1– Feb 2014
- Xiangli Yang, Zixing Song, Irwin King, A Survey on Deep Semi-supervised Learning
- Subhabrata Mukherjee, Sentiment Analysis (2012), https://www.researchgate.net/publication/236203597_Sentiment_Analysis_A_Literature_Survey/link/00b7d51c746a1c5aab000000/download