**Capstone Proposal**

**1.Sentiment Analysis**

Sentiment analysis is one of the fundamental tasks of natural language processing. It is commonly applied to process customer reviews, implement recommender systems and to understand sentiments in social media.

**Problem Statement**

The task involves the binary classification of tweets(Twitter data) from twitter as expressing either positive or negative sentiment.

**Datasets**

Link: <https://www.kaggle.com/c/twitter-sentiment-analysis2/data>

File descriptions

train.csv - the training set  
test.csv - the test set  
Data fields

ItemID - id of twit  
Sentiment - sentiment  
SentimentText - text of the twit

0 - negative

1 - positive

**Solution Statement**

The solution to this problem is to provide a binary sentiment classification (0 for negative, 1 for positive) for each item in the testing data using feature vectors from a self-trained model.

**Execution Plan**

steps:-1 Prepossessing of tweets.

1.1 NoUrl

1.2 NonNumeric

1.3 Tokens

1.4 Stemming

1.5 Lemma

1.6 Stopwords

1.7 Removing words less then 3 char.

Step2:- spliting data into train test.

Step3:- Choosing a right algorithms or different different algorithm with high accuracy.

Step4: train model based on training set and then predict on test set.