Topic 1

Sentiment Analysis

Sentiment analysis is the method which classifies the text into positive, negative, and neutral. It is extensively used by the e-commerce website to get the idea about how the product is doing in the market. It is also used in the election in predicting the wave of any candidate. It is also called opinion mining.

We have used VADER (valence aware dictionary sentiment reasoning) technique for analysing the tweet. We have used the dataset of the tweets on the farmer’s protest. It is applicable also on unlabelled text.

Firstly, we have imported all the necessary library which will be used. Also, we have imported the dataset.

We will also download the important package from the nltk library like stopwords, punkt, wordnet. Stopwords are the word which are used to connect the sentence and doesn’t add any emotion.

We have read the dataset using the pandas. Now we can access the rows and column of the tweet content.

After we have tokenize the sentence, this means that we have divided the sentence in each word. It is present the punkt package.

After that we have imported SentimentIntensityAnalyzer() from the nltk. It gives the polarity scores of positive, negative, neutral, and compound. We have then iterated it for the whole row of content.

This is good and easy technique, but it is not considering the interrelation between the word it only searches for only one word.

Eg.: happy=positive

Not happy=negative

Sad=negative

Not sad=positive,when any negation words comes in the sentence it reverses it sentiment.

First the sentence is tokenized into words. Bag of word or count vectorisation technique is in which the text is converted numeric system. And then sentence is combined to form a vector. Now the cosine similarity is used here.

It needs the labelled data one part for training purpose and other part for testing.

Tools used:-

1. Numpy
2. Pandas
3. NLTK

Project link –

https://colab.research.google.com/drive/1igT4Ivd8BcJneefADd\_ip1xpjvsVlZCx?usp=sharing