READ ME

**Motivations**

Being extremely interested in everything having a relation with the Machine Learning, the independent project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can make estimations, predictions and give the ability for machines to learn by themselves is both powerful and limitless in term of application possibilities. We can use Machine Learning in Finance, Medicine, almost everywhere. That’s why I decided to conduct my project around the Machine Learning.

**Idea**

This project was motivated by my desire to investigate the sentiment analysis field of machine learning since it allows to approach natural language processing which is a very hot topic actually.

**The Project**

Sentiment analysis, also refers as opinion mining, is a sub machine learning task where we want to determine which is the general sentiment of a given document. Using machine learning techniques and natural language processing we can extract the subjective information of a document and try to classify it according to its polarity such as positive, neutral or negative. It is a really useful analysis since we could possibly determine the overall opinion about a selling object, or predict stock markets for a given company like, if most people think positive about it, possibly its stock markets will increase, and so on. Sentiment analysis is actually far from to be solved since the language is very complex (objectivity/subjectivity, negation, vocabulary, grammar, but it is also why it is very interesting to working on.

In this project we choose to try to classify tweets from Twitter into “positive” or “negative” sentiment by building a model based on probabilities. Twitter is a microblogging website where people can share their feelings quickly and spontaneously by sending a tweets limited by 140 characters. You can directly address a tweet to someone by adding the target sign “@” or participate to a topic by adding an hastag “#” to your tweet. Because of the usage of Twitter, it is a perfect source of data to determine the current overall opinion about anything.

Also, to add on we did a YouTube comments analysis into “positive”, “negative” and “neutral” comments with their percentages depicted in a pie chart. To add on we used web scrapping to fetch the top trending news and happenings.

modules used:

PYTHON

* nltk
* nltk.corpus import stopwords
* nltk.tokenize import word\_tokenize
* nltk.classify.util as util
* nltk.classify import NaiveBayesClassifier
* nltk.metrics import BigramAssocMeasures
* nltk.collocations import BigramCollocationFinder as BCF
* itertools
* pickle
* os.path
* statistics import mode
* nltk.classify import ClassifierI
* nltk.metrics import BigramAssocMeasures
* nltk.collocations import BigramCollocationFinder as BCF
* sys

graphical user interface:

**nodejs modules**

express

bootstrap

Passport/passport-local/passport-local-mongoose

MongoDB

**webscraping modules:**

MongoDB

Cheerio.js

Request-Promise

**RESULTS**

Software was able to fetch tweets from tweeter and predict positive and negative on it

Secondly top trending hash tags were identified which when clicked show public sentiment over them

Thirdly sentiment analysis on YouTube comment was successfully implemented

Fourthly voice input was given as complementary feature.

Diagrammatical view(in form of pie chart was also given to user) of sentiment analysis also provided