Interim Project Report

IIT Kanpur - Summer Internship 2020
Applied Machine Learning and Data Science 2020
Course Code - 002

Project: Sentimental Analysis of Tweets using Python
3 July, 2020

Project Members

1. **Kaustubh Dwivedi**
   (kaustubhdwivedi1729@gmail.com)
2. **Tanmay Misra**
   (tanmaymisra619@gmail.com)

Overview:

What is sentiment analysis?
Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. Sentiment analysis allows businesses to identify customer sentiment toward products, brands or services in online conversations and feedback.

Project Motivation:

Why Twitter Sentiment Analysis:
Sentiment Analysis Dataset Twitter has a number of applications:

Business: Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customers' feelings towards products or brand, how people respond to their campaigns or product launches and also why consumers are not buying certain products.

Politics: In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analyzing election results.

Public Actions: Twitter Sentiment Analysis also is used for monitoring and analyzing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere.

Project Aims and Objectives:

Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc), and even on intentions (e.g. interested v. not interested). Main aim of this project include collecting various tweets and feed backs and then classifying them into their proper categories positive, neutral or happy and also detecting emotions, like happiness, frustration, anger, sadness, and so on.

Technologies:

This project is of Natural language processing (NLP) which is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language. It is based on Python technology and uses many inbuilt libraries that helps in creating it.
Major Libraries used: numpy, panda, seaborn, matplotlib, sklearn, nltk, wordcloud, genism, tqdm, etc.

Difficulties of Sentimental Analysis:

Sentiment analysis can be applied to many areas but arriving at whether a statement is positive or negative can be difficult. The categorization is mainly split into two types: facts and opinions.Facts are expressed about entities, whereas opinions are about their properties.

Furthermore, opinions are completely subjective and describe people's sentiments, appraisals or general feelings towards entities and their properties.

The human language can be complex for machine-based learning systems to interpret. For example, opinions can be expressed with sarcasm or irony, and the order of words can add even more confusion.

Other Problems Faced: Irony and sarcasm, Types of negations, Word ambiguity, Multipolarity, etc

Further Application and Domains:

Sentiment analysis can also be used to:
1. Analyze tweets and/or facebook posts over a period of time to detect sentiment of a particular audience
2. Monitor social media mentions of your brand and automatically categorize by urgency
3. Automatically route social media mentions to team members best fit to respond
4. Automate any or all of these processes
5. Gain deep insights into what's happening across your social media channels