Twitter sentiment analysis using Machine Learning and Natural Language Processing techniques

Aayushi Somani, Amal Krishna A, Ruturaj Chothe, Poornima Deshpande, Vishaal Sharma, Kamalnath Irrukulla, Uttam Lakhani, Shubham Pawar, Rushikesh Khot, Herat Makwana, Ishank Jain, Anjali Shah, Aishwarya Salimath

**Abstract**

Social media analysis is the process of collecting data from popular social networking services and predicting the public view on any given domain based on the analysis of the collected data. This is achieved using machine learning and natural language processing techniques, along with various python libraries such as matplotlib, tweepy and textblob. The sentiment analysis tool has a simple user interface, which asks for a keyword based on which analyses of the tweets containing the keyword are segregated and statistically represented in terms of the sentiments being expressed as positive, negative or neutral.

The first step towards training the machine learning model begins with collecting the datasets that are made available using the Twitter API, which is achieved using the Tweepy library. It offers labelled datasets that can be used to efficiently train the machine learning model. The tweet processing and classification is done using the textblob library, which offers a simple API for natural language processing tasks such as sentiment analysis and classification.

With this text classifier, we can label each Tweet as positive, negative or neutral of the sentimental value in a few minutes. However, human language is complex. Teaching a machine to analyze the various grammatical nuances along with diverse cultural variations, slang and misspellings that occur in social media make the process complex and teaching a machine to understand how the context affects the tone proves to be a challenging area. Humans are fairly intuitive when it comes to interpreting the tone of a piece of writing.

Sentiment analysis has its own limitations like any other ML predictive and is not used as a 100% accurate marker but with a little supervision it can be a great asset.