**International Journal of Synthetic Emotions**
Volume 11 • Issue 1 • January-June 2020

# Real Time Sentiment Analysis

Sandip Palit, Academy of Technology, Kolkata, India
Soumadip Ghosh, Academy of Technology, Kolkata, India
https://orcid.org/0000-0003-4817-5363

## ABSTRACT

Data is the most valuable resource. We have a lot of unstructured data generated by the social media giants Twitter, Facebook, and Google. Unfortunately, analytics on unstructured data cannot be performed. As the availability of the internet became easier, people started using social media platforms as the primary medium for sharing their opinions. Every day, millions of opinions from different parts of the world are posted on Twitter. The primary goal of Twitter is to let people share their opinion with a big audience. So, if the authors can effectively analyse the tweets, valuable information can be gained. Storing these opinions in a structured manner and then using that to analyse people's reactions and perceptions about buying a product or a service is a very vital step for any corporate firm. Sentiment analysis aims to analyse and discover the sentiments behind opinions of various people on different subjects like commercial products, politics, and daily societal issues. This research has developed a model to determine the polarity of a keyword in real time.

## KEYWORDS

Acronyms, Cursor, Emojis, NLTK Corpus, Polarity, Tokens, Twitter API, Web 2.0

## 1. INTRODUCTION

Sentiment analysis is the process of analysing people's way of thinking, and feelings towards a particular product or service. Before taking any decision, we try to gather other's opinion on that topic (Shayaa et al., 2018). Previously we used to ask people, for their personal opinion. But with the advent of Web 2.0, social media became the primary platform for sharing our opinions. So, analysing the posts and tweets became the better option than personally collecting the feedbacks. The main steps of sentiment analysis are (Guevara et al., 2018): Data acquisition, Text processing, Feature extraction, Sentiment classification, Evaluation and Results. There are various approaches for sentiment analysis, like the Machine learning approach and the lexicon-based approach. In the machine learning based approach, the model was trained using a labelled dataset. The lexicon-based approach depends on the dictionary or a bag of words containing pre-tagged lexicons. Tools for sentiment analysis include python NLTK, GATE, Opinion finder, LingPipe and LIWC (Alessia et al., 2018). Brand monitoring, customer service, market research and analysis are some of the applications of sentiment analysis. Brandwatch reported that every second, six thousand tweets are posted on Twitter (van Dijck J., 2011). So, it is the best source for analysing opinions. Twitter API allows us to extract the tweets within a

DOI: 10.4018/IJSE.2020010103

rate limit. TextBlob is a python library for Natural Language Processing tasks, like translation and language detection, sentiment analysis, tokenization and part-of-speech tagging. Natural Language Toolkit (NLTK) provides the stop words, which is important for feature extraction.

## 2. LITERATURE REVIEW

Web 2.0 (Murugesan S., 2007) is an enhanced version of the Web 1.0. It forms the foundation for social media platforms. It is strongly characterised by the change from static to dynamic or user-generated content. Some of its advantages are: better media support, dynamic and real-time discussion. It enables the user to add their opinions in the form of posts or tweets. Web 2.0 tools (Thackeray et al., 2008) allows the user to create and modify content on many social platforms, like Twitter, Facebook and Youtube. This promotes interactive content, which results in a better user experience. A major part of this data is unstructured texts, such as tweets, reviews and blogs.

Although the first academic studies for analysing public opinion was during World War 2, the evolution of modern sentiment analysis took place in the mid-2000s, whose main purpose was to understand people's opinion on various online products (Kumar & Vadlamani, 2015). In recent years, researchers started applying sentiment analysis on social media platforms like Twitter and Facebook. This also works well on various other topics like the stock market, disasters, medicines, election and software engineering (Mäntylä et al., 2018).

'Opinion mining and sentiment analysis' by Pang and Lee (2008) was the top-cited paper on sentiment analysis. It focuses mainly on the fundamentals and basic application of Sentiment analysis. It also developed some free resources like lexicons and datasets.

One of the pioneer works on Reviews analysis done by Pang, Lee and Vaithyanathan (2002), tried to classify the overall statement, instead of classifying by topics. Using standard machine learning techniques like Naïve Bayes, Support Vector Machine and Maximum Entropy Classification, they classified the movie reviews as positive or negative.

Turney's works (2002) on document level semantic classification was also a widely cited work from 2002. He developed a simple unsupervised learning algorithm to classify reviews as thumbs up (recommended) or thumbs down (not recommended), based on the semantic orientation of the phrases present in the review. They achieved an average accuracy of 74%.

Mamta and Ela Kumar (2019) developed a lexicon-based framework to perform real-time sentiment analysis on Twitter. In the data pre-processing stage, they removed the special characters from the tweets, indirectly removing the emojis. Emojis are important in understanding the positive or negative tone of the statement (Walther & D'addario, 2001). In our approach, we manually decoded these emojis. Nowadays, we use lots of acronyms. Usage of acronyms can decrease the accuracy of our model (Palmquist R.D., 2008), so we expanded those acronyms using the python regex module.

## 3. APPROACHES FOR SENTIMENT ANALYSIS

Sentiment analysis techniques can be classified mainly into three main categories: the Lexicon based approach, the Machine learning approach and the Hybrid approach (Dorothy & Rajini, 2016).

### 3.1. Machine Learning Based Approach

Machine learning approaches can be classified into supervised, unsupervised and semi-supervised learning techniques. In sentiment analysis, mainly the supervised learning technique is followed. It requires a large number of labelled training documents. The model is trained based on the previous dataset, and then that model is used to classify an unseen text. The accuracy of this technique is greatly influenced by the dataset.