International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Sentiment Analysis using Machine Learning

Simran Patil[1], Arif Shaikh[2], Sakshi Singh[3]

[1-3]*Information Technology, Mumbai University, Theem College of Engineering, India*

Abstract: Thanks to tremendous expanse of opinions of users, their reviews, feedbacks and suggestions on the market over the online resource, it's most indispensable to find, analyze and consolidate their views for increased deciding. Our aim is to produce a sentiment analysis system for client review classification, which will be useful to analyze the data within the style of number of tweets wherever opinions are extremely unstructured and either positive or negative. Sentiment analysis may be a machine learning tool that analyzes texts for polarity, from positive to negative. By coaching machine learning tools with samples of emotions in text, machines mechanically find out how to discover sentiment without any human input method of deciding the emotional tone behind a series of words, want to gain associate understanding of the attitudes, opinions and emotions expressed among an internet mention. The aim of this project is to make a sentiment analysis model which can enable to categorise words supportive of their sentiments, that is whether they are positive, negative and additionally the magnitude of it.

Keywords: Twitter, text mining, polarity, sentiment analysis.

i. Introduction:

Sentiment analysis (or opinion mining) may be a tongue process technique to verify whether knowledge is positive, negative or neutral. Sentiment analysis is usually performed on matter knowledge to assist businesses, monitor whole and products sentiment in client feedback, and perceive what client wants. Sentiment analysis is the method of sleuthing positive or negative sentiment in text. It is usually utilized by businesses to discover sentiment in social knowledge, gauge whole name, and perceive customers. Since customers are categorical about their thoughts and feelings additional overly than ever before, sentiment analysis is turning into a necessary tool to observe and perceive that sentiment. Mechanically, analyzing client feedback, like opinions in survey responses and social media conversations, permits brands to be told what makes customers happy or pissed off, so they will tailor product and services to satisfy their customers' wants. Sentiment analysis models concentrate on polarity (positive, negative, neutral) however additionally on feelings and emotions (angry, happy, sad,etc), urgency (urgent, not urgent) and even intentions (interested, v. Not interested).

ii. Methodology:

In the coaching method, our model learns to associate a selected input (i.e. a text) to the corresponding output (tag) supported, taking a look at the samples used for coaching. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. positive, negative or neutral) are fed into the machine learning rule to get a model.

In the prediction method, the feature extractor is employed to rework unseen text inputs into feature vectors. These feature vectors are then fed into the model, that generates foreseen tags (again, positive, negative or neutral).

iii. Motivation:

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover, the response on twitter is

more prompt and also more general. Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms an also estimate how well this responding in the market, which areas of the market is it having a favourable response and in which a negative response. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

iv. Literature Review:

Sentiment analysis of in the domain of micro-blogging is a relatively new research topic so there is still a lot of room for further research in this area. Decent amount of related prior work has been done on sentiment analysis of user reviews, documents, web blogs/articles and general phrase level sentiment analysis [5]. These differ from twitter mainly because of the limit of 140 characters per tweet which forces the user to express opinion compressed in very short text. The best results reached in sentiment classification use supervised learning techniques such as Naive Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised (e.g. [3] and [4] and semi-supervised (e.g. [1] and [2]) approaches, and there is a lot of room of improvement. Various researchers testing new features and classification techniques often just compare their results to base-line performance. There is a need of proper and formal comparisons between these results arrived through different features and classification techniques in order to select the best features and most efficient classification techniques for particular applications.

v. Algorithms Implemented:

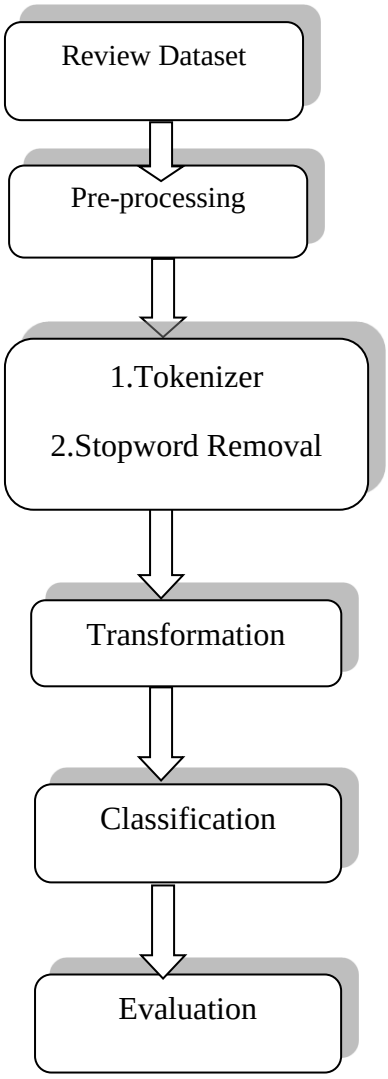
Naive Bayes:
Naive Bayes is a fairly straightforward cluster of probabilistic algorithms that, for sentiment analysis classification, assigns likelihood that given a word or phrase ought to be thought-about positive, negative or neutral. Naive Bayes calculates words against one another. So, with every machine learning models trained for word polarity, we are able to calculate the chance that a word, phrase or text is positive, negative or neutral.

When techniques like lemmatization, stopword and removal, and TF-IDF are enforced with Naive Bayes and become a lot of predictively correct.

Linear Regression:
Linear regression is a statistical algorithm rule, used if we want to predict a Y value, given X features. Using machine learning, the datasets area unit examined to point out a relationship. The relationships area unit are then placed on the X/Y axis, with a line running through them to predict additional relationships.

Support Vector Machines:
A Support Vector Machine is another supervised learning machine learning model, like simple regression, however additionally advanced. SVM uses algorithms to coach and classify text inside our sentiment polarity model, taking it a step on the far side X/Y prediction.

vi. Process:



vii. Future Work:

viii. Conclusion:

Overall, the success and failures of all these different approaches gave us a good overall picture of the challenges of sentiment analysis, and provide some guidelines for sentiment analysis with other sets of data in the future. First, we note the use of colloquial and slang language in most of the confessions. The use of spell checking corrected for this somewhat. Nonetheless, the synset and sentiment lexicons we used are better suited to more formal styles of writing. An alternative approach is to replace our synsets and lexicons with "slang" versions or even the automatic generation of sentiment lexicons on a slang corpus.

Another area of interest is the difficulty in correlating topics with sentiment. Intuition says that topics themselves should portray different sentiments, and so should be useful for sentiment analysis. This method turns out be fairly crude, as sometimes topics may be too neutral or too general to actually be good indicators of mood. It is possible for someone to complain angrily about their current relationship, or laugh because of a happy moment during the relationship. All of these get mapped into the same topic, but each has a substantially different mood.

ix. Acknowledgements:

The authors would like to thank Assistant Professor Sonali Karthik for their guidance.

x. References:

1. Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In proceedings of international conference on Language Resources and Evalutation (LREC), 2010.
2. Luciano Barbosa and Junlan Feng. Robust Sentiment Detection on Twitter from Biased and Noisy Data. In proceedings of the international conference on Computational Linguistics (COLING), 2010.
3. Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classifications of Reviews. In proceedings of the Annual Meeting of the Association of Computation Linguistics (ACL), 2002.
4. Samuel Brody and Nicholas Diakopoulus. Using Word Lengthening to Detect Sentiment in Microblogs. In proceedings of Emperical Methods on Natural Language Processing (EMNLP), 2011.

5. Theresa Wilson, Janyce Wiebe and Paul Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In the Annual Meeting of Association of Computational Linguistics: Human Language Technologies (ACL-HLT), 2005.