COMPUTOLOGY: Journal of Applied Computer Science and Intelligent Technologies
Volume 1, Issue 2, Jul-Dec 2021, pp. 37-46
DOI: 10.17492/computology.v1i2.2105
https://www.journalpressindia.com/cjacsit

Twitter Sentiment Analysis using Machine Learning

Prerna Agarwal and Pranav Shrivastava***

ABSTRACT

Twitter Sentiment Analysis is a vital field of natural language processing (NLP) and data analytics that focuses on gauging public opinion, emotions, and attitudes expressed in tweets. It involves the use of various NLP techniques, machine learning algorithms, and data mining methods to classify tweets as positive, negative, or neutral. Researchers and organizations employ sentiment analysis to understand customer feedback, track brand perception, predict market trends, and monitor public sentiment during events or crises. It's a powerful tool for social listening, enabling businesses, policymakers, and individuals to make informed decisions and engage with their audience effectively in the fast-paced and dynamic world of Twitter.

Keywords: Sentiment Analysis; Twitter, Learning algorithms; Natural Language processing.

1.0 Introduction

Twitter allows businesses to engage personally with consumers. However, there's so much data on Twitter hard for brands to prioritize which tweets or mentions to respond to first. That's why sentiment analysis has become a key instrument in social media marketing strategies. Sentiment analysis is a tool that automatically monitors emotions in conversations on social media platforms.

Twitter has become an exponential growth in the use of social media to share feelings, thoughts, and ideas gold culture to analyze brand performance. The ideas found on Twitter are casual, honest, and informative. What can be selected from more formal surveys, etc. Millions of about brands to interact with. These feelings, if identified, can be beneficial to companies simply by observing them identify the brand, aspects, and time 7 periods that evoke a sense of polarization. These are brands products, celebrities, events, or political parties. Thus, the teams were devoted to analysis of the candidate's performance during public opinion polls or public reaction to the launch of the gadget or movies. But with more than 500 million tweets a day, that information is already large enough analyzed manually by any team. Similarly, the variety of tweets cannot be determined[1][2].

The set of rules is prepared by hand. Note that the task is to understand the mood in the tweet more complex than any well-formatted document. Tweets do not follow any official language structure and they contain words taken from the official language (i.e., words outside the dictionary). Often used to express emotions (phrases, expressions, etc.)[3]. Especially we are building a computational model that can classify a given tweet as positive, negative, or neutral reflects the mood. There will be polar tweets expressing the mood classes.

**Corresponding author; JEMTEC, Greater Noida, NCR, India (E-mail: prerna115@gmail.com)*
***GL Bajaj Institute of Technology, Greater Noida, NCR, India (pranav.paddy@gmil.com)*

However, a neutral class can include an objective or subjective tweet or reflect user neutrality there are no thoughts or ideas. Examples of each class can be found in Table 1. The solution is to use three classes are conducted considering to the current situation research in this area. Experiences that informally predict our feelings show that they are our system this work[4]. We also create an interactive 16 visualization tool to help companies interpret using our emotional predictor to imagine public opinion about products and brands.

Table 1: Example of Class

Class	Tweet
Positive	@hon1paris: I <3 1D too! #muchlove
Negative	The new Transformers suck!! Wasted my time and money!!!
Neutral	Well, I guess the govt did what it could. More needed though! I plan to wake up early in the morning #early2bad

Our discussion of the methodology for text classification has been set up this way: we will show how to establish the model and explain the basics of how text is classified. We then summarized the results of the literature review to understand and define the area of study[5]. Once gaps are identified, system details are adjusted. The best function for this system is to provide a reliable and robust system. This system is made up of a combination of functions that provide a reliable and robust system. We then experiment with this system for comparison.

The performance of our system with other published studies. In addition, the effectiveness of each added function the system is highlighted. Finally, we summarize our research contributions by looking at what our next steps would be. The second part discusses our procedure for creating an interactive rendering tool using a sensitivity forecast model. We summarize the results of the literature review to understand what happened metrics and functions worked well and how they can be adapted to visualize sentiment analysis[6]. We describe other relevant general tools that perform a similar task. Highlighting past works, describe our approach and system architecture for developing a visualization tool. We further suggest screenshots app explaining the motivation of each view and user element. Finally, we summarize our contribution and the steps to be taken in the future.

2.0 Objective

Estimation examination of within the domain of micro-blogging may be a moderately unused investigate subject so there is still a parcel of room for assist investigate in this range. Not too bad sum of related opinion examination of client surveys, archives, web blogs/articles and common express level estimation examination. These contrast from twitter basically since of the limit of 140 characters per tweet which forces the client to precise conclusion compressed in exceptionally brief content. The leading comes about reached in opinion classification utilize administered learning methods such as Credulous Bayes and Support Vector Machines, but the manual naming required for the directed approach is exceptionally expensive. I found a piece of light I found a piece of love we will forever be in the depths of the dark and the cold the pieces of one. Different analysts testing modern highlights and classification techniques regularly fair compare their comes about to base-line execution. There’s a require of legitimate and formal comparisons between these comes about arrived through distinctive highlights and classification techniques in order to choose the most excellent highlights and most productive classification strategies for particular applications.

3.0 Scope

Sentiment analysis of within the domain of micro-blogging may be a moderately unused investigate subject so there is still a parcel of room for encourage investigate in this zone. Better than average sum of related assumption examination of client audits, archives, web blogs/articles and common express level estimation examination. These contrast from twitter basically since of the limit of 140 characters per tweet which forces the client to precise supposition compressed in exceptionally brief content. The leading comes about reached in assumption classification utilize administered learning methods such as Credulous Bayes and Support Vector Machines, but the manual naming required for the administered approach is exceptionally expensive. Unsupervised and semi-supervised approaches, and there is a part of room of change. Different analysts testing modern highlights and classification techniques regularly fair compare their comes about to base-line execution. There’s a require of appropriate and formal comparisons between these comes about arrived through diverse highlights and classification techniques in order to choose the leading highlights and most productive classification procedures for particular applications.

4.0 Literature Survey

Assumption investigation may be a developing zone of Common Dialect Preparing with investigate extending from document level classification [7] to learning the extremity of words and expressions (e.g.[8])Given the character limitations on tweets, classifying the estimation of Twitter messages is most comparable to sentence- level assumption investigation (e.g., ([9]; [10])); be that as it may, the casual and specialized dialect utilized in tweets, exceptionally nature of the micro blogging space make Twitter opinion investigation a really distinctive assignment. It’s an open address how well the highlights and methods utilized on more well-formed information will exchange to the microblogging space.

Other analysts have started to investigate the utilize of part-of-speech highlights but comes about stay blended. Highlights common to micro blogging (e.g., emoticons) are moreover common, but there has been small investigation into the convenience of existing opinion assets created on non-micro blogging data.

Analysts have moreover started to explore different ways of consequently collecting preparing information. Several analysts depend on emoticons for characterizing their preparing information (Pak and Paroubek 2010; Bifet and Straight to the point 2010). (Barbosa and Feng 2010) misuse existing Twitter assumption destinations for collecting preparing information. (Davidov, Tsur, and Rappoport 2010) too utilize hashtags for making training information, but they constrain their tests to sentiment/non-sentiment classification, instead of 3-way extremity classification, as we do.

5.0 Details about Twitter

Sentiment analysis is a widely used method of text production. Analyzing emotions on Twitter therefore means using advanced to analyze the sentiments of the text (here, tweet) in positive, negative, and neutral forms. Also known as Opinion Mining, it is primarily used for conversations, opinions, and exchanges to set business strategies, political analysis, and to evaluate public action. Polarity detection is the most popular method of sentiment analysis, which involves categorizing statements as Positive, Negative, or Neutral.

Figure 1 Graph 1

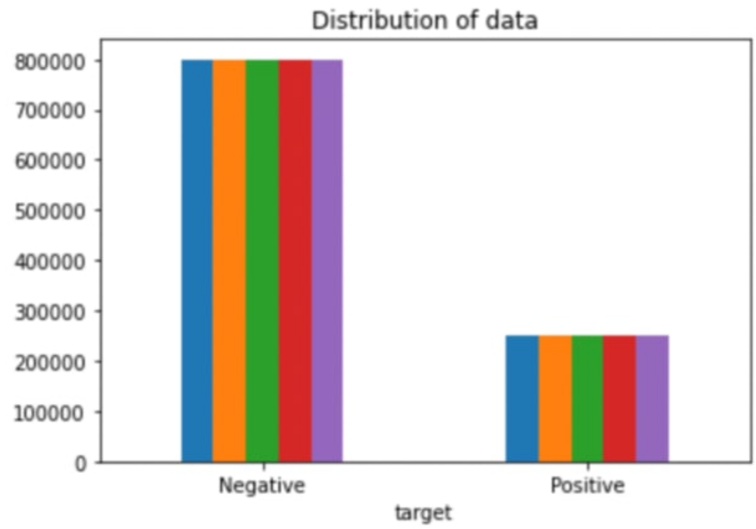
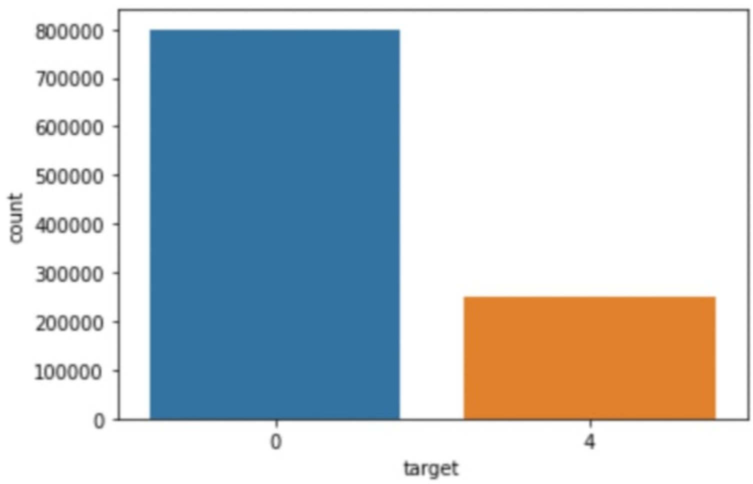


Figure 2 Graph 2



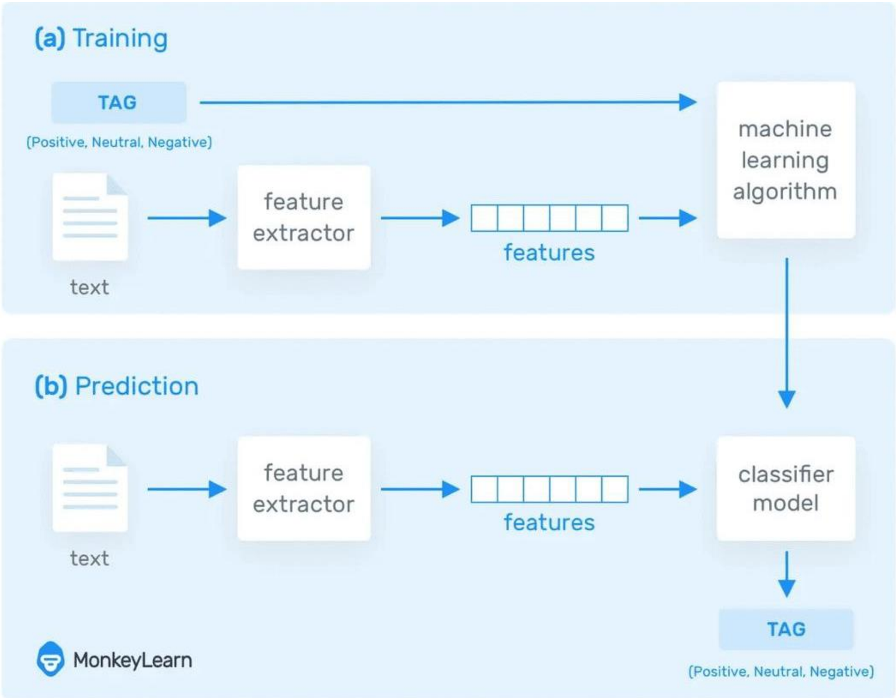
Before training the model, we did numerous pre-processing procedures on the dataset, mostly deleting stop words and emojis, as stated in the issue statement. For better generalization, the text document is subsequently transformed to lowercase. Following that, the punctuations were cleaned and eliminated, minimizing the amount of noise in the dataset. After that, we eliminated the repetitive characters from the words URLs, as they were of little use. Finally, for better results, we performed Stemming (reducing words to their derived stems) and Lemmatization (reducing derived words to their base form known as lemma)[11].

This data can be used allocation, improve organizational performance, provide better products/services, and, ultimately, improve citizen lifestyles and human connections in order to establish a better society. The influence of measuring people's feelings on products, services, and events, for example, allows enterprise managers to have knowledge and parameters to make decisions. Another example is city council administrators, who may have the potential to improve citizen services and more effectively manage growth and sustainability concerns based on community feedback.

The contemporary environment for data collecting and analysis of people's sentiments is social media. People can share and discuss anything, from personal ideas to current events and societal problems. Access to social media can potentially reveal additional data hidden metadata. Language of the operating system, device kind, capture time, and geographic location are only a few examples. In prior research, the main usefulness of sentiment analysis was determining if the conveyed view in the document or sentence was favorable, negative, or neutral[12]. However, it was useless in decision-making because no reason for the shift in sentiment was recognized. As a result, a mechanism for analyzing public opinion fluctuations was required.

Despite the potential benefits, there are some drawbacks to using automatic analysis, including the difficulty of implementing it due to the ambiguity of natural language It's a beautiful, sunny day, and characteristics provided information hashtags, emoticons, and links, making it challenging to decipher the expressed emotion. Furthermore, automatic procedures are required, which necessitate big datasets of annotated posts or lexical databases with emotional terms correlated with sentiment scores. Another key feature is that analyses are appropriate for English, which is a limitation for other languages.

Figure 3: Working of Sentiment Analysis



In the subject of sentiment analysis, there are various obstacles in terms of design and application domains with ambiguous or scarce datasets in a variety of circumstances. In addition, there is a scarcity of labeled data, which can stymie progress in this field.

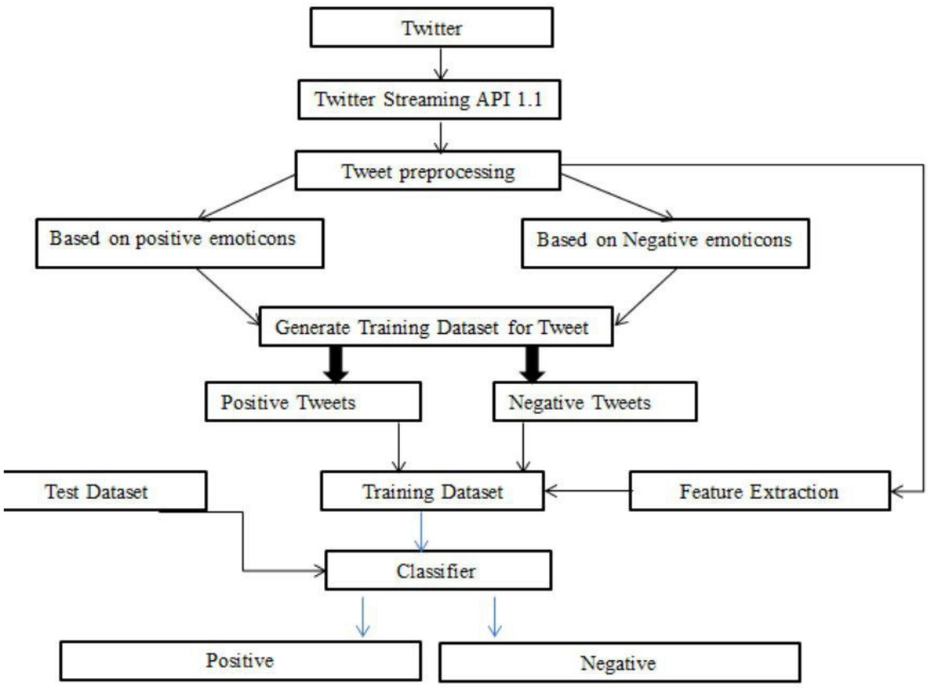
Sentiment analysis is the strategy utilized for understanding people's feelings and sentiments, with the assistance of machine learning, with respect to a specific item or benefit. Opinion investigation models require a tall volume of a specific dataset. One of the foremost challenging perspectives of creating and preparing a demonstrate is obtaining the proper volume and sort of opinion investigation dataset. At upgrade, we have compiled a list of ten open datasets that can assist you get begun along with your extend on assumption examination.

There are a number of techniques and complex algorithms used to control and train machines to analyze emotions. This article will discuss some of the most commonly used algorithms but they can give great results when used.

Online reputation most valuable assets for brands. Social media misconceptions can be costly for a company if they are not addressed effectively and quickly. By analyzing emotions on Twitter, you can track what is being said on service and help identify unhappy customers or negative comments before they escalate. However, analyzing feelings on Twitter can provide valuable insight to help you make decisions. What do customers like about your brand? Which aspects are highlighted more negatively? Aspect Emotion Analysis using Twitter can show you which aspects of your business need to be improved and what sets you apart from the competition.

Being available on Twitter is becoming increasingly important for customer service agents. Six out of ten users expect the brand to answer each customer's question within an hour. But how do you determine which help requests are most important? Emotion analysis on Twitter allows you analyze all the interactions between your brand and your customers, so you can be sure that you are responding to the most pressing issues in the first place[13].

Figure 4: Classification of Emotion



Twitter is the main source of information for consumers. In fact, people use it to express different feelings, observations, beliefs, and opinions on different topics. You can use Twitter emotion analysis to track specific keywords and topics to identify customer trends and interests. If you are planning to launch a new product, it is important to understand what customers like, how their behavior changes over time.

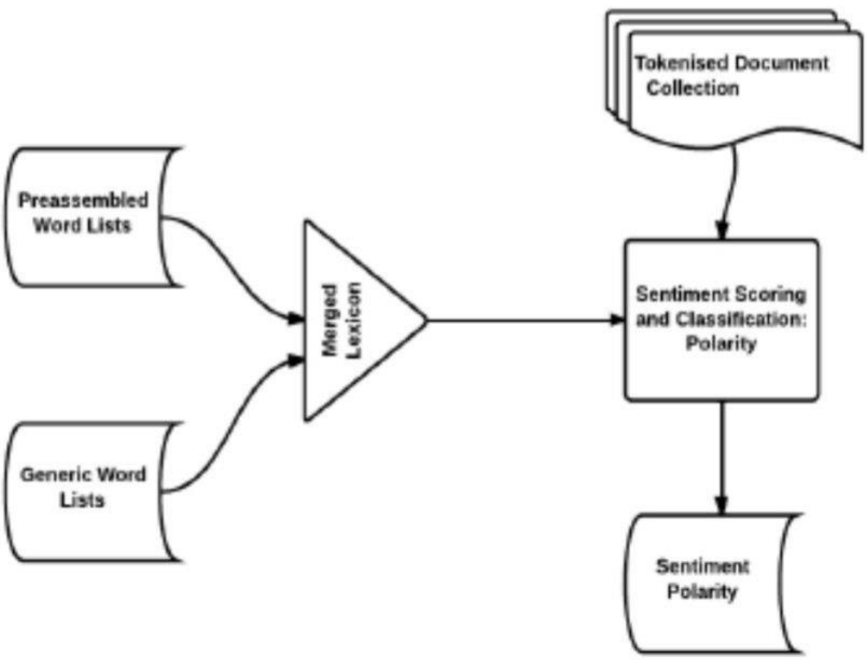
Below is an example of how Twitter used emotion analysis to track 4,000 tweets called halal food. This information allowed researchers to identify different reasons for consuming halal food and to divide their market into different types of consumers .Analyzing Twitter emotions will help you stay one step ahead of your competitors. By identifying the pain points of competitors, you can focus on these areas when promoting your business.

The pre-processed data set has many unique features the way we separate attributes, we separate the sides processed data set. This aspect is then used to calculate positive and negative polarization in a sentence that is useful for determine people's opinions using models such as unigram, bigram. Machine learning methods require a basic presentation properties of the text or documents being processed. This follows property vectors for the classification problem.

6.0 Lexicon Attitudes

The Scale the thesaurus save money and subscribe to the information, please share polarization I value emotions, words, statements, positive statements, negative and objective dictionary stored in the dictionary. Commonly used substances in sentiments, lexical content vocabulary, viz. stock trading and sales fees traditional terms, phrases, and new idioms for traditions Genre type as well as lexicon search memory.

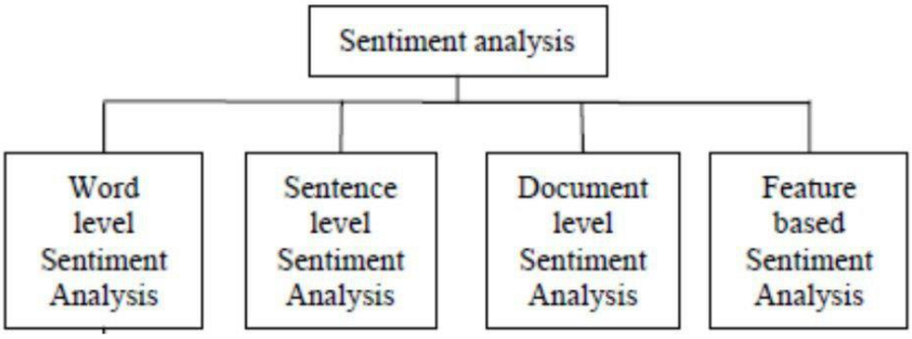
Figure 5 Lexicon Model



7.0 Levels of Sentiment Analysis

Tasks on paradox section can be done at several levels of granularity.

Figure 6: Levels of Sentiment Analysis



7.1 Document level

Marks their individual documents at the document level, the entire document is classified positive or negative class.

Basic approach: Find mood poles in certain sentences or words and combine them to find the polarity document. Other approaches: Complex linguistic phenomena, such as solving connected connections, pragmatic, etc.

- However, there are several functions:
- Objective: To classify the emotions of the whole document
- Classes: positive, negative and neutral
- Assumption: Each document is focused on one object.

(wrong in discussions, blogs, etc.) and include single thought.

7.2 Sentence or phrase level

Sentiment-level sentiment analysis is associated with tags individual sentences with mental poles. Speech level classification Divides a sentence into categories positive, negative or neutral degree.

Basic approach: Find the sentimental direction of individual words in sentences / phrases and then combine them to identifythe tonality of the whole sentence or phrase. Other approaches: consider the discursive structure of the text

However, there are several functions:

- Task 1: Identify subjective / objective sentences

Lessons: objective and subjective

- Task 2: Sentimental classification of sentences

Classes: positive and negative

Assumption: The sentence consists of only one an idea that may not always be true

7.3 Aspect level of Feature level

It’s also about naming each word with a mood determine the nature of voting. Problems with classifying emotions by aspect or character level identify and extract product features from the source information. Techniques such as dependency analyst and discussion uses constructions.

However, there are several functions:

- Objective 1: Identify and extract the properties of an existing object comments made by the thinker (e.g. commentator)

- Objective 2: Identify ideas about functions negative, positive or neutral
- Task 3. Find synonyms for characters.

7.4 Word level

Polarity and prepositions in later works a phrase to classify emotions in speech and document levels In classifying the sense of the word, adjectives such as are mainly used features but additions, There are two ways to automatically record emotions word level:

(1) Dictionary-based approaches

(2) The corps is approaching

8.0 Conclusion

Data analysis based on micro blogs includes: all located in areas of development and division. I'm trying to come up with an idea personnel, according to the name of the commune, have the same results in the building, and most important presentation. We work with each other on all models of models; you can download these models, add more information, copywriter, slow down and verbal you can put a window in front of the window and the result of the outputs may be different. If you find a model in this window, click on it. We have a very slow learning curve large and small in the language of the auxiliary polarity of the exchange field polarization. For example, if you are a slow-moving user, you can do this easily create give it a different touch than the word, see more minimize the ifs effect if desired.

Moreover, we are currently focusing only on unigrams and their effects bigrams and trigrams can be studied.However, we need more to make big features and trigrams an effective function marked the data set as our 9,000 tweets.Now we are studying certain parts of speech using the unigram samples, we may try to include POS data. Slight decrease in performance) with a significant decrease in accuracy only adjectives are used as signs. But for these results comments can be checked for classification and analysis of emotions in micro blogs sites like Twitter. Another feature worth investigating is whether there is information about it the relative position of a word in a tweet has some effect on how the classifier works. While Pang et al. have studied a similar function and reported negative results on their own the results were based on comments that were very different from tweets and they worked in a very simple model.

One of the possible problems in our study three classes not equal. The objective class, which includes 4,543 tweets, is about twice as large positive and negative classes, including 2543 and 1877 tweets, respectively. The problem with unequal classes is that the classifier tends to improve overall accuracy most systems by increasing class accuracy, even if this happens the cost of reducing the accuracy of minority classes. That's why we positive or significantly higher objective level of accuracy report negative classes. To eliminate this problem and so that the classifier does not indicate that any class is incorrect, more data (tweets) should be marked so that all three are present our classes are almost equal.

We go at a general analysis emotions there are the ability to work analysis emotions with a partially specific context typically use our website for certain types of keywords they can be divided into several separate classes, namely: politicians / politicians, celebrities, products / brands, sports / athletes, media / movies / music. To separate analysis of the sense of tweets belonging to only one classes (i.e., curriculum information is specific to one of them, not general category) and if we apply a general analysis of emotions to it, compare the results obtained instead of.

Finally, model a person's self-confidence systems. For example, if we have 5 people tagging each tweet, we draw tweet in a two-dimensional plan of objectivity / subjectivity and positivity / negativity the difference of all 5 tag matching tweets, only 4 agree, only 3 agree whether or not most votes agree. We can develop our cost function come up with optimized class boundaries that give more weight all 5 tags are agreed and those tweets with the transactions discounts, as well as certain weights. Thus, the influence of human faith can be imagined in the analysis of emotions.

References

1. Albert Biffet and Eibe Frank. Sentiment Knowledge Discovery in Twitter Streaming Data. Discovery Science, Lecture Notes in Computer Science, 2010, Volume 6332/2010, 1-15, DOI: 10.1007/978-3-642-16184-1_1
2. Alec Go, Richa Bhayani and Lei Huang. Twitter Sentiment Classification using Distant Supervision. Project Technical Report, Stanford University, 2009.
3. Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010.
4. Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner and Isabell M. Welpe. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2010.
5. Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.
6. Chenhao Tan, Lilian Lee, Jie Tang, Long Jiang, Ming Zhou and Ping Li. User Level Sentiment Analysis Incorporating Social Networks. In Proceedings of ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), 2011.
7. Efthymios Kouloumpis, Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.
8. Hatzivassiloglou, V., & McKeown, K.R.. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, 2009.
9. Johann Bollen, Alberto Pepe and Huina Mao. Modelling Public Mood and Emotion: Twitter Sentiment and socio-economic phenomena. In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.
10. 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.
11. Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL), 2002.
12. Rudy Prabowo and Mike Thelwall. Sentiment Analysis: A Combined Approach. Journal of Infometrics, Volume 3, Issue 2, April 2009, Pages 143-157, 2009.