



REVIEW MONITORING BY SENTIMENTAL ANALYSIS

1st Rajanala Siva Bagya Jyothi
Parul institute of Engineering and Technology
Parul University Vadodara, India

2nd Rajanala Venkata Sravani
Parul institute of Engineering and Technology
Parul University
Vadodara, India

3rd Emani Jagadish Reddy
Parul institute of Engineering and Technology
Parul University
Vadodara, India

4th kasu Mahesh Vardhan Reddy
Parul institute of Engineering and Technology
Parul University
Vadodara, India

5th Dr. Sunita Yadwad
Associate Professor
Parul institute of Engineering and Technology Parul University
Vadodara, India

I. ABSTRACT

With the advancement of web technology and its growth, there is a huge volume of data present in the web for internet users and a lot of data is generated too. Internet has become a platform for online learning, exchanging ideas and sharing opinions. Social networking sites like Twitter, Facebook, Google+ are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. There has been lot of work in the field of sentiment analysis of twitter data. This survey focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases.

II. INTRODUCTION

Due the penetration of the internet in all domains of life which has led to increase of people's participation actively and give remarks as an issue of communicating their concern/feedback/opinion in various online forums. Although most of the times these comments are helpful for the creator to extemporize the substance that is being provided to people, but sometimes these may be abusive and create hatred-feeling among the people. Thus as these are openly available to the public which is being viewed from various sections of the society, people in different age groups, different communities and different socio-economic background, it becomes the prime responsibility of the content creator (the host) to filter out these comments in order to stop the spread of negativity or hatred within people. Lately there has been many cases in which the growing menace of hate and negativity has been witnessed in the online platforms especially social media as such, many governments around the world has seen the rise of cases related to cyber bullying that has led to spread of hatred and violence. Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information. Sentiment analysis is the process of classifying whether a block of text is positive, negative, or, neutral. Sentiment analysis is contextual mining of words which indicates the social sentiment of a brand and also helps the business to determine whether the product which they are manufacturing is going to make a demand in the market or not. The goal which Sentiment analysis tries to gain is to analyze people's opinion in a way that it can help the businesses expand. It focuses not only on polarity (positive, negative neutral) but also on emotions (happy, sad, angry, etc.). It uses various Natural Language Processing algorithms such as Rule-based, Automatic, and Hybrid.

There are major challenges in sentiment analysis approach:

- I. If the data is in the form of a tone, then it becomes really difficult to detect whether the comment is pessimist or optimist.
- II. If the data is in the form of emoji, then you need to detect whether it is good or bad.
- III. Even the ironic, sarcastic, comparing comments detection is really hard.
- IV. Comparing a neutral statement is a big task.

In the statistical context, Machine Learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While Machine Learning involves concepts of automation, it requires human guidance. Machine Learning involves a high level of generalization in order to get a system that performs well on yet unseen data instances. Machine learning is a relatively new discipline within Computer Science that provides a collection of data analysis techniques. Some of these techniques are based on well established statistical methods (e.g. logistic regression

and principal component analysis) while many others are not.

Most statistical techniques follow the paradigm of determining a particular probabilistic model that best describes observed data among a class of related models. Similarly, most machine learning techniques are designed to find models that best fit data (i.e. they solve certain optimization problems), except that these machine learning models are no longer restricted to probabilistic ones. NLP Projects referred as Natural Language processing. We develop NLP projects which works based on machine to understand human speech, activity and reply human understandable format.

We design NLP

academic Projects as human language given as input called natural language and given to computer to analyze human language and transformed to useful representation form. Our main aim to implement NLP projects for students are to obtain valuable information from human speech enabled by computer and to understand what user need it

III. LITERATURE REVIEW

Writing Audit:

Survey Checking by Estimation Investigation

Assumption examination, too known as supposition mining, has ended up a capable apparatus for businesses to get it client input through online audits. This survey writing digs into the application of opinion examination for survey observing, investigating its benefits, strategies, and existing challenges.

Benefits of Audit Observing with Estimation Investigation:

Pick up Client Experiences:

Estimation examination makes a difference categorize surveys as positive, negative, or impartial, giving a speedy get a handle on of client opinion towards items, administrations, or brands.

Recognize Ranges for Enhancement:

By pinpointing negative surveys, businesses can recognize viewpoints requiring change, such as item highlights or client benefit.

Track Brand Notoriety:

Checking assumption over time permits businesses to gage the adequacy of promoting campaigns and track their generally brand notoriety.

Prioritize Client Issues:

Assumption investigation can highlight repeating negative concerns, empowering businesses to prioritize client benefit endeavors.

Competitive Examination:

Businesses can use estimation investigation to compare client recognition of their brand against competitors.

Strategies for Survey Estimation Investigation:

Lexicon-Based Approach:

This strategy utilizes pre-defined word references of positive and negative words to decide opinion.

Machine Learning Methods:

Directed machine learning calculations are prepared on labeled datasets to classify opinion. Well known methods incorporate Credulous Bayes and Bolster Vector Machines (SVM).

Profound Learning Procedures:

Profound neural systems, such as Repetitive Neural Systems (RNNs), are progressively utilized for estimation investigation due to their capacity to handle complex dialect structures.

Challenges of Survey Assumption Investigation:

Mockery and Incongruity:

Assumption investigation instruments may battle to decipher mockery or incongruity, driving to misclassifications.

Domain-Specific Dialect:

Audits frequently utilize casual dialect or industry language that will not be captured by common opinion vocabularies.

Invalidation Taking care of:

Invalidation words like "not" can switch the assumption of a sentence, requiring vigorous refutation dealing with procedures.

Spam and Fake Audits:

The nearness of spam or fake audits can skew opinion examination comes about.

Future Bearings:

Investigate in assumption investigation is continually advancing. Future patterns incorporate:

Aspect-Based Opinion Examination:

This method goes past fundamental opinion classification and recognizes particular viewpoints of a item or benefit that clients are positive or negative approximately.

Opinion Investigation over Stages:

Coordination opinion investigation over different online stages, like social media and audit destinations, can give a more all encompassing see of client opinion.

Multilingual Estimation Examination:

As businesses work all inclusive, estimation examination apparatuses that can handle different dialects will ended up progressively critical.

IV. METHODOLOGY

Data in the form of raw tweets is acquired by using the Python library “tweepy” which provides a package for simple twitter streaming API . This API allows two modes of accessing tweets: SampleStream and

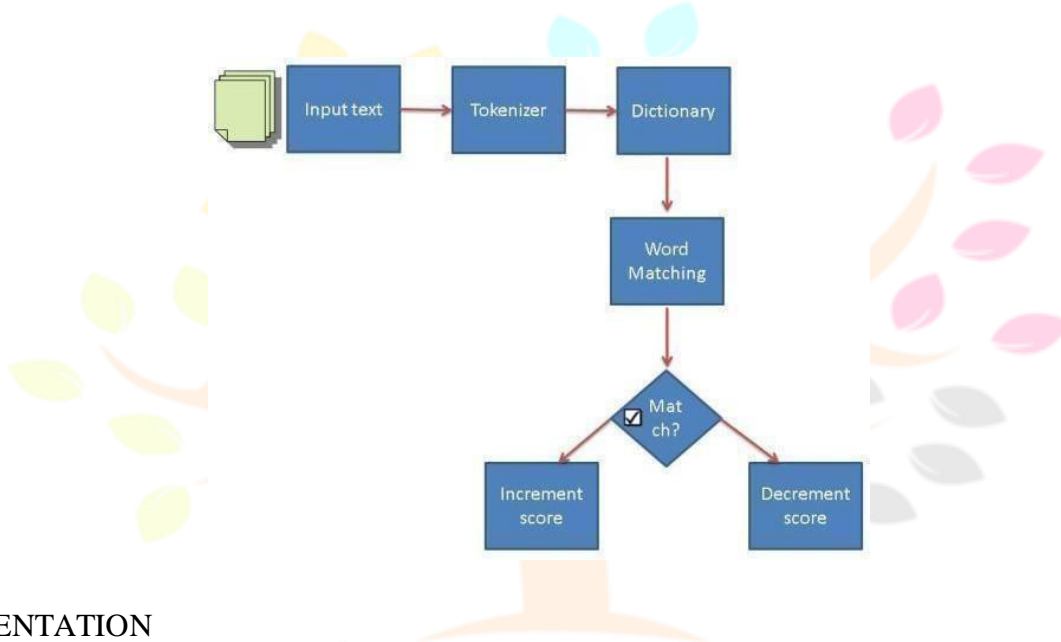
FilterStream. SampleStream simply delivers a small, random sample of all the tweets streaming at a real time.

FilterStream delivers tweet which match a certain criteria.

It can filter the delivered tweets according to three criteria:

- Specific keyword to track/search for in the tweets
- Specific Twitter user according to their name
- Tweets originating from specific location(s) (only for geo-tagged tweets).

A tweet acquired by this method has a lot of raw information in it which we may or may not find useful for our particular application. It comes in the form of the python “dictionary” data type with various key-value pairs.



C. IMPLEMENTATION

```

[25]: print("checking polarity for each tweet")
def polarity(text):
    testimonial = TextBlob(text)
    polarity = testimonial.sentiment.polarity
    return polarity

def subjectivity(text):
    testimonial = TextBlob(text)
    subjectivity = testimonial.subjectivity
    return subjectivity

def senti(text, polarity_threshold=0.2):
    testimonial = TextBlob(text)
    senti = testimonial.sentiment.polarity

    if senti >= polarity_threshold:
        return 'Positive'
    elif np.abs(senti) < polarity_threshold:
        return 'Neutral'
    else:
        return 'Negative'

tweets_data['polarity'] = tweets_data['text_nlp'].apply(lambda x: polarity(x))
tweets_data['subjectivity'] = tweets_data['text_nlp'].apply(lambda x: subjectivity(x))
tweets_data['sentiment'] = tweets_data['text_nlp'].apply(lambda x: senti(x))
tweets_data.head()
  
```

```
[4]: tweets_data.isnull().sum()
[4]: id      0
      date   0
      user_name 1
      text   0
      hashtags 0
      dtype: int64

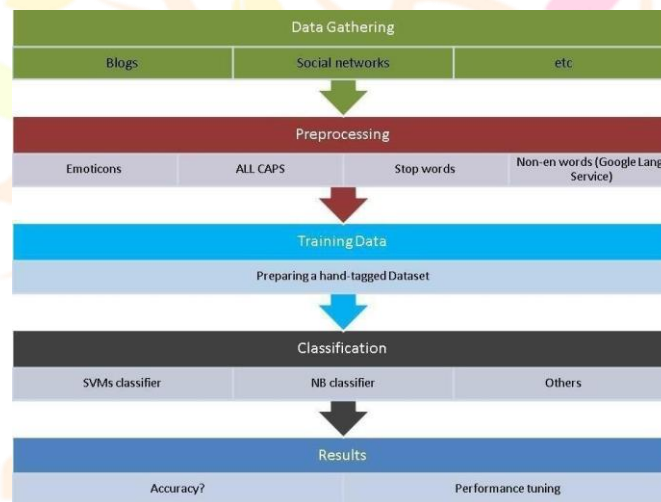
[5]: tweets_data.replace(to_replace=["\\t|\\n|\\r|\\t|\\n|\\r"],value="",regex=True,inplace=True)
      print("escape sequence removed")
      import numpy as np
      tweets_data.replace("^[^$]*$",np.nan,regex=True,inplace=True)
      tweets_data.dropna(inplace=True)
      tweets_data

escape sequence removed

[5]:
```

	id	date	user_name	text	hashtags
0	1.340540e+18	20-12-2020 6:06	Rachel Rich	Some folks said daikon paste could treat a cy...	PfizerBioNTech
1	1.338160e+18	13-12-2020 16:27	Albert Fong	While the world has been on the wrong side of ...	Notags
2	1.337860e+18	12-12-2020 20:33	ellreu	#coronavirus #SputnikV #AstraZeneca #PfizerBio... coronavirus.SputnikV.AstraZeneca.PfizerBioNTech...	
3	1.337860e+18	12-12-2020 20:23	Charles Adler	Facts are immutable. Senator: even when you're...	Notags
4	1.337850e+18	12-12-2020 20:17	Citizen News Channel	Explain to me again why we need a vaccine @Bor...	whereareallthesickpeople PfizerBioNTech
...
197865	1.440560e+18	22-09-2021 6:30	VarBLR	18-44 #URBAN #Bengaluru #CovidVaccine Availabl...	URBAN,Bengaluru,CovidVaccine,COVISHIELD
197866	1.440560e+18	22-09-2021 6:09	CoWIN Blore 18-44	Pincode:500066SPUTNIK V - Dose 1: 25 slotsAge...	Notags
197867	1.440560e+18	22-09-2021 6:00	VarBLR	45+ #URBAN #Bengaluru #CovidVaccine Availabl...	URBAN,Bengaluru,CovidVaccine,COVISHIELD
197868	1.440560e+18	22-09-2021 5:59	CoWIN Blore 18-44	Pincode:560070SPUTNIK V - Dose 2: 20 slotsAge...	Notags
197869	1.440550e+18	22-09-2021 5:46	NewsBytes	Sputnik V is another vaccine being used in Ind...	Notags

D. WORK FLOW



V. CONCLUSION AND FUTURE WORK

Right now we have worked with only the very simplest unigram models; we can improve those models by adding extra information like closeness of the word with a negation word. We could specify a window prior to the word (a window could for example be of 2 or 3 words) under consideration and the effect of negation may be incorporated into the model if it lies within that window. For example if the negation is right next to the word, it may simply reverse the polarity of that word and farther the negation is from the word the more minimized its effect should be. Apart from this, we are currently only focusing on unigrams and the effect of bigrams and trigrams may be explored. As reported in the literature review section when bigrams are used along with unigrams this usually enhances performance. However for bigrams and trigrams to be an effective feature we need a much more labeled data set than our meager 9,000 tweets

We tried to build a Toxic comment classification with the help of sentiment analysis system by studying and implementing algorithms of machine learning. We implemented Naive Bayes and Maximum Entropy algorithms. Baseline model performed the worst with no doubt as it had least number of features. The modular system we've built can easily be scaled for new algorithms be it in Machine Learning, Deep learning or Natural Language Processing. Sentiment analysis system is an active field of research and we can still further improve our system by working more on the algorithms, trying out different things in preprocessing and checking which ones get the best precision metrics.

VI. REFERENCE

- i. P. Vidyullatha¹, Satya Narayan Padhy¹, Javvaji Geetha Priya², Kakarlapudi Srija³, Sri Satyanjani Koppiseti⁴, "Identification and Classification of Toxic Comment Using Machine Learning Methods", IVth year B.Tech Student, Dept. of Computer Science and Engineering,
- ii. Koneru Lakshmaiah Education Foundation, A.P., India, 2021
- iii. Pallam Ravi, Hari Narayana Batta, Greeshma S, Shaik Yaseen, "Toxic Comment Classification", Anurag Group of Institutions, Telangana, India, may-jun 2019
- iv. Zhongkai Hu (huzhongkai@zju.edu.cn), Jianqing Hu (qhu@zju.edu.cn), Weifeng Ding (vkkharry@gmail.com), Xiaolin Zheng (xlzheng@zju.edu.cn), "Review Sentiment Analysis Based on Deep Learning", College of Computer Science Zhejiang University Hangzhou, China, 2015
- v. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Aristidis G. Vrahatis, Vassilis P. Plagianakos, "Convolutional Neural Networks for Toxic Comment Classification", Department of Computer Science and Biomedical Informatics, University of Thessaly Lamia, Greece, 2018
- vi. S. Rahamat Basha, T. Bhasara Reddy, "Imbalanced Text Features for Toxic Comments Classification", Department of Computer Science and Technology, Sri Krishnadevaraya University, India, January 2022
- vii. Ahmad Alsharif,¹ Karan Aggarwal,² Sonia,¹ Deepika Koundal,³ Hashem Alyami,⁴ and Darine Ameyed⁵
¹ Yogananda School of Artificial Intelligence, Computing and Data Science, Shoolini University, Solan, Himachal Pradesh 173229, India
² Electronics and Communication Engineering Department, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala 133207, India
³ Department of Systemics, School of Computer Science, University of Petroleum Energy Studies, Dehradun, India
⁴ Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
⁵ System Engineering Department, Ecole de Technologie Supérieure, University of Quebec, Montreal, Canada, "An Automated Toxicity Classification on Social Media Using LSTM and Word Embedding", 2022
- viii. Zhigang Xu, Kai Dong, Honglei Zhu*, "Text sentiment analysis method based on attention word vector", School of Computer and Communication, Lanzhou University of Technology, (2020)
- ix. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of

Twitter Data”, Department of Computer Science Columbia University New York, NY 10027 USA,(2011)

- x.Zhixue Zhao zhixue.zhao@sheffield.ac.uk, Ziqi Zhang ziqi.zhang@sheffield.ac.uk, Frank Hopfgartner f.hopfgartner@sheffield.ac.uk, “A Comparative Study of Using Pre-trained Language Models for Toxic Comment Classification”, University of Sheffield Sheffield, UK, 2021
- xi.Hassan Saifa , Yulan Heb , Miriam Fernandez , Harith Alania, “Contextual Semantics for Sentiment Analysis of Twitter”, aKnowledge Media Institute, Open University, United Kingdom bSchool of Engineering and Applied Science, Aston University, United Kingdom, (2016)
- xii.Sara Zaheri, Jeff Leath, David Stroud, “Toxic Comment Classification”, Southern Methodist University, 6425 Boaz Lane, Dallas, Texas 75205 szaheri, jleath, jdstroud @smu.edu (2020)
- xiii.Zulfadzli Drus, Haliyana Khalid*, “Sentiment Analysis in Social Media and Its Application: Systematic Literature Review”, Azman Hashim International Business School, Kuala Lumpur, 54100, Malaysia (2019)
- xiv.Babak Abedin 1, Abdul Babar 2, Alireza Abbasi 3, “Characterization of the Use of Social Media in Natural Disasters: A Systematic Review” 1 2 Faculty of Engineering and IT University of Technology Sydney Sydney, Australia Babak.Abedin@uts.edu.au Abdul.Babar@uts.edu.au, 3 School of Engineering and IT The University ofNew South Wales Canberra, Australia Alireza.Abbasi@unsw.edu.au, (2014)
- xv.zhigang Xu, kai Dong, honglei Zhu*, “Text sentiment analysis method based on attention word vector”, School of Computer and Communication, Lanzhou University of Technology,(2020)



International Research Journal
IJNRD
Research Through Innovation