International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-10 Issue-4, November 2021

Sentiment Analysis-Enhancements and Applications

Aayush Gupta, Anant Gandhi, Saarthak Agarwal, Shamin Chokshi, Saravanakumar Kandasamy

Abstract: ***The concept of Natural Language Processing that deals with problems of identifying the sentiment from the voice or text of a speaker or writer and then use that analysis further for making predictions, market survey, customer service, product satisfaction, precision targeting etc. is called Sentiment analysis. From one viewpoint, it is an abstract evaluation of something dependent on close to home observational experience. It Is mostly established in target realities and incompletely governed by feelings. Then again, a sentiment can be deciphered as a kind of measurement in the information in regards to a specific subject. It is a lot of markers that mix present a point of view, i.e., perspective for the specific issue. So as to enhance the accuracy of sentiment analysis/classification, it is imperative to appropriately recognize the semantic connections between the various words and phrases that are describing the subject or aspect. This can be done by applying semantic analysis with a syntactic parser and supposition vocabulary. This research will discuss different sets of approaches for application or domain specific problems and then compare them to obtain the best possible approaches to the problem of sentiment analysis.***

Keywords: Sentiment Analysis, Lexicon, Support vector machine, Regression

I. INTRODUCTION

There are loads of genuine circumstances in which sentiment analysis may demonstrate helpful. In Marketing, to know how people in general respond to your item/administration, which parts of it they like/disdain, how they need it to be improved, and so forth [5][10]. In Quality Assurance, to identify blunders in your items dependent on your genuine clients experience. In Politics, to decide the perspectives on the individuals with respect to explicit circumstances, what are they furious or cheerful for, and so forth. In Risk Prevention, to recognize if a few people are being assaulted or badgering (for instance in Social Media or inside visits). In Finances, to attempt to recognize the feeling towards a brand, to envision their market moves [5][10][11].

Manuscript received on September 09, 2021.
Revised Manuscript received on September 16, 2021.
Manuscript published on September 30, 2021.
*Correspondence Author
Aayush Gupta, School of Computer Science and Engineering, VIT Vellore, India, E-mail. aayush.gupta2017@vitstudent.ac.in
Anant Gandhi, School of Computer Science and Engineering, VIT Vellore, India, E-mail. anant.gandhi2017@vitstudent.ac.in
Saarthak Agarwal, School of Computer Science and Engineering, VIT Vellore, India, E-mail. Saarthak.2018@vitstudent.ac.in
Shamin Chokshi*, School of Computer Science and Engineering, VIT Vellore, India, E-mail. Shamin.2018@vitstudent.ac.in
Saravanakumar K., School of Information Technology and Engineering, VIT Vellore, India, E-mail. ksaravanakumar@vit.ac.in

Individuals share information, encounters and considerations with the world by utilizing Social Media like websites, discussions, wikis, survey destinations, informal organizations, tweets, etc. This has changed the way in which individuals convey and impact social, political and financial conduct of others [5][12]. The underlying foundations of sentiment analysis/classifications lie in the examinations on general opinion investigation towards the start of twentieth century and in the content subjectivity examination performed by the Natural language and computer based semantics network in 1990's.[2] Notwithstanding, the episode of computation based assumption examination just happened with the accessibility of abstract messages on the Web. Thus, 99% of the researches have been distributed after 2004. [2] Sentiment Analysis researches are dissipated to numerous distribution settings, and the consolidated number of studies in the best 15 scenes just portray ca. 30% of the studies altogether. The work of Minqing Hu and Bing Liu circa 2004 [1] really set the ball rolling in terms of sentiment analysis research as we know it today, at least in the customer reviews case. From the early twentieth century, started the process of documenting a general public opinion by gathering all sorts of data in different forms, primarily questionnaires. A formal scientific journal on this was put together in the year 1937[2] [4]. Sentiment analysis employs multiple NLP (Natural Language Processing) algorithms and approaches. Major types of these algorithms are: rule-based, automatic, and hybrid. Typically, a rule-based framework utilizes a lot of human-made guidelines to help recognize subjectivity, extremity, or the subject of a supposition [9] [10]. These principles may incorporate different systems created in computational phonetics, for example, Stemming, tokenization, grammatical feature labelling and parsing; Lexicons (for example arrangements of words and articulations). Rule-based frameworks are extremely naive since they don't consider how words are consolidated in an arrangement [13][9][10]. Obviously, further developed preparing methods can be utilized, and new standards added to help new articulations and jargon. In any case, including new rules may influence past outcomes, and the entire framework can get extremely mind boggling. Since rule-based frameworks frequently[6][7][8] require calibrating and upkeep, they'll likewise require ordinary speculations [13]. Automatic approaches, unlike rule-based frameworks, don't depend on physically made standards, however on AI/ML systems. A sentiment analysis task is typically displayed as an order issue, in which a classifier bolsters a content and formulates a classification, for example positive, negative, or neutral [10][14].

The initial phase in an AI/ML content classifier is to change the content extraction or content vectorization, and the old style approach/algorithm has been bag-of-words or bag-of-n grams with their recurrence. All the more as of late, new element extraction procedures have been applied dependent on word embeddings (otherwise called word vectors). The classification of text generally includes a statistical machine learning model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks [14]. Naïve Bayes is a family of probability based algorithms which employ Bayes's Theorem to foresee the category of a text [15]. Linear Regression is an extremely popular algorithm in statistical mathematics and is used to approximately calculate some value (Y) given a set of features (X) [16] [13]. Support Vector Machines is a non-probabilistic approach that employs a graph plot of text/content examples as points in a multidimensional space. Examples of multiple classifications (sentiments) are mapped to differentiated areas inside that space. Then, testing and training text data is classified into a category based on how it is similar to existing text data and the areas they're mapped to [17]. Deep Learning is a diverse set of algorithmic approaches that attempt to stimulate the human brain, and uses artificial neural networks for data processing. Hybrid models and approaches integrate the most prominent characteristics of rule-based and automatic techniques/algorithms into one model/approach [18]. A vital advantage of these models is that outcomes have a fairly larger chance of being progressively precise. This research will discuss in detail the approaches and algorithms that were followed for sentiment classification and categorization and how some of these approaches prove extremely useful and efficient for some vital real world applications of sentiment analysis. 12 of some very important works in the field will be elaborately defined and summarized here. Since these multiple studies will have different evaluation methods to measure and compare performances of algorithms, the most reliable ones will be chosen to describe and to compare approaches. However, the studies might discuss the approach for differing applications which makes it difficult to clearly distinguish which technique should be the most optimal as it is very subjective matter. Based on the results of these evaluation methodologies, all base models and their respective approaches will have a detailed comparison. Finally our research will conclude by extracting and discussing the possible inferences from the results of comparison. Future improvements or directions (if any) will be defined by the end of our discussion. Section II will define terminologies and frequently used words, common equations etc. that are essential for the reader to understand the study. Section III will contain the individual discussion of base methods, their approaches and their applications. This will be followed by the defining of evaluation methods which will be used to assess the performance of these algorithms which will be done in Section IV. The following section, Section V will compare performances of discussed algorithms in tabular format using different parameters. Section VI will conclude the research with all findings, inferences and finally future directions. Section VII, the last segment will have all references and all citations that have been used.

II. CRUCIAL TERMINOLOGIES

Sentiment analysis -The process of digitally or programmable identifying and categorizing opinions expressed during a piece of text, especially so as to work out whether the writer's attitude or feelings towards a specific topic, product, situation etc. is positive, negative, or neutral. Lexicon- Lexicon is that the total stock of words, paragraphs, phrases, conversations and word elements that carry meaning. (Linguistics) The a part of the grammar that include the lexical entries* for all the words within the language and which can also include various other information counting on particular theory of grammar. User trust mining -Identifying and realizing patterns in several users reviews on topics, products ,situations etc and linking them to at least one another during a graph and identifying whether which persons opinion will another person trust Regression -Regression may be a statistical measurement utilized in finance, investing, AI , machine learning and other disciplines that attempts to work out the strength of the connection between one variable (usually denoted by Y) and a series of other changing variables Genetic algorithm- A genetic algorithm may be a search heuristic that's inspired by Charles Darwin's theory of natural evolution. This algorithm reflects the method of survival where the fittest individuals are selected for copy so as to supply offspring of subsequent generation. Big data - Extremely large data sets which will be analysed computationally to reveal patterns, trends, and associations, especially concerning human behaviour and interactions.
Sentiment classification -Sentiment classification is that the task of watching a bit of text and telling if someone likes or dislikes the thing they're talking about. The input X may be a piece of text and therefore the output Y is that the sentiment which we would like to predict. Sentiment similarity- Sentiment similarity indicates the similarity between two words from their underlying sentiments by usage of varied algorithms of machine learning of AI .(this approach is employed to mine trust of users in e-commerce websites.) Support vector machine- Support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and multivariate analysis . BERT- Bidirectional Encoder Representations from Transformers (BERT) may be a technique for NLP (Natural Language Processing) pre-training developed by Google. BERT was created and published in 2018 by Jacob Devlin and his colleagues from Google. Google is leveraging BERT to raised understand user searches. Learning Algorithm -A learning algorithm may be a method wont to process data to extract patterns appropriate for application during a new situation. especially , the goal is to adapt a system to a selected input-output transformation task.

III. FRAMEWORK USED FOR SENTIMENT ANALYSIS

The first step to analyse a public opinion is scraping the data from the internet whether its collecting movie reviews,

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

public opinion on social media, product reviews, hotel reviews etc. The above comes under data collection. A suitable API can be made in order to collect the data. Then the procedure of preprocessing extracted records for analysis is a key pillar inside the sentiment analysis process, it consists of reducing noise opinion and effect .Other related responsibilities such as feature extraction and sentiment class without compromising the word meaning[14]. From a syntactic point of view, extricated information commonly incorporates additional words, called "commotion", which builds capacities zone measurement. examined the effect of preprocessing onclass calculations, tests affirmed that stop phrases, rehashed letters, numbers and URLs evacuation minimally affect classifiers, and in addition, it could upgrade widely calculation assets. The preprocessing methodologies productivity relies intently upon stop words realities units and abbreviations lexicons selection[24].Three essential preprocessing systems: -

- Grammatical feature (POS) labeling, which iseffective for express highlights extraction, however needs execution when literarysubstance contains understood ones.
- Stemming and Lemmatization: Stemming examine with the transformation of an expression to its root structure with out setting thought, while Lemmatization is setting cognizant and rely on extra word reference measurements to improve exactness.
- Stop Word Removal: words with high recurrence inside the printed content and do never again incorporates significant realities for the assessment.

Grouping is the ensuing principle challenge for estimation assessment, all the past advances were improved to allow a superior precision and execution in making sense of the extremity of text[25][26]. Classifiers normally offer limitsin conditions, for example, mockery, incongruity, utilization of slang, and so forth anyway with the huge canvases on preprocessing techniques, the difficulties changed allowing specialists to plunge into prevalent sort systems and investigate better approaches to play out this crucial.

Usually there are short bodies of textual content will come with a label such as tremendous or bad sentiment. The first method to building a class version from words isto use the bag of phrases/unigrams version. In order to convert texts into bag of words, first you increase a vector of all words appearing the entire set of text inside the schooling set. Next, you iterate through each textual content in the training set[15].The characteristic set will become extremely big and sparse. In order to reduce the function length to boom computation velocity and performance of category models, several feature selection techniques have been used on the bag of words models. The bag-of-phrases model is one among the best language fashions used in NLP[17]. It makes an unigram version of the text by retaining track of the variety of occurrences of every word. This can later be used as a features for Text Classifiers. In this bag-of-words model you handiest take character words into account and give every phrase a specific subjectivity score. It is simple to implement however can affect the accuracy[42][43]. The sentence “This e-book isn't always good” could be interpreted as a superb sentence, unless any such assemble is implemented. Another instance is that the sentences “This e-book may be very good” and “This book is good” may have the equal rating with a unigram model of the textual content, however now not with an unigram + bigram model[53][54]. The weighted words are then put through variousalgorithms for semantic as well as sentiment analysis. There are various hybrid multi layer algorithms that are researched for highly accurate sentiment analysis with high precision recall. The results of the proposed algorithm is evaluated on measures of accuracy , F1-score, polarity score and precision recall which indicates us the superiority of the algorithm[24]. Usually hybrid frameworks take a good amount of time to train the neural network than the simple algorithms but it also means that they can be cable to provide high accuracy and quality of feedback on the input[22][23].

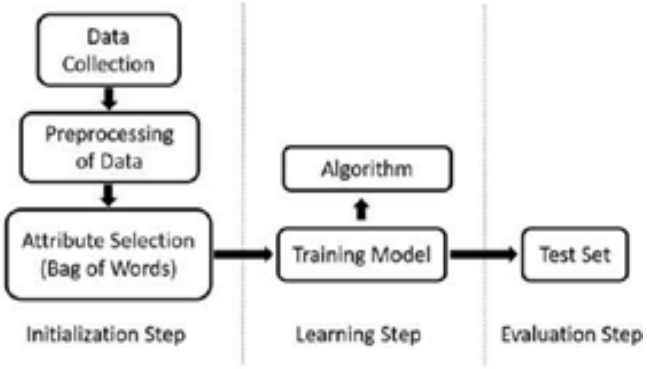

Fig -1 Sentiment Analysis Framework

IV. LITERATURE REVIEW

This section discusses the research under two subtitles –

- Approach
- Applications.

The approach part deals with the discussion of the approach and enhancement algorithm used in the studies. And the Application p a r t discusses the various applications in which the proposed approach can be applied to.

1) Sentiment analysis of short text data with a bi levelattention model [19]

Approach

This model proposes a brief text sentiment analysis methodology supported bi-level attention model, that doesn't trust manual options or external information. Uncommonly, at word level, our modelimproves the effect of word delineation by bringing idle theme information into word-level etymology representation. Neural point model is utilized to locate the idle subject of the content. a substitution theme word consideration component is presented to investigate the phonetics of words from the edge of point word relationship; At the arrangement level, an auxiliary consideration instrument is utilized to catch the connection among local and worldwide conclusion articulation. Analyses on the ChnSentiCorp-Htl-ba-10000 and NLPCC-ECGC datasets approve the adequacy of the BAM model. Our model joins neural subject model and a center system to moderate the effect of data exiguity.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

At word level, our model improves the effect of word outline by bringing idle theme information into word-level semantics portrayal. a substitution theme word consideration component is anticipated to investigate the semantics of words from the point of subject. At grouping level, consideration system is utilized to catch the connection among local and universal slant articulation. Investigations on the ChnSentiCorp-Htl-ba-10000 and NLPCC-ECGC datasets approve the viability of the BAM model. Blunder examination shows that the model stays helpless against disarray once positive and negative choices appear inside a similar book. this sort of blunder originated from the inability to get a handle on the need of local options[31]

Application

This methodology is best for short content information. Short content information is scanty which implies the it is dissipated or scattered, for opinion examination this causes an issue as it needs co-event of words which influences word embeddings that rely upon co event .Another issue is that less legitimate highlights of short content limit the models capacity to remove content highlights and feeling analysis[32]. To help this outside information is utilized to adapt to the above issues yet outer information requires different excellent manual highlights and subsequently this technique is restricted then again settling sparsity issue by dimensionality decrease requires high computational unpredictability. Information sparsity to sum things up printed content experiences continually been a difficulty in feeling arrangement Existing techniques are constrained in the extent of utility extension on the grounds that few manual highlights are required or rely upon top notch outside data base in explicit fields Moreover, a few procedures goal to cure sparsity issue with the guide of applying dimensionality decrease, for example, topic designs. Subject models are utilized to find the relationship among words and idle points, even as inert theme added substances are utilized to reduce the dimensionality of the content Furthermore. Be that as it may, the heft of the related research has concentrated on the situation of point information on content characterization, without thinking about the relationship among theme and opinion.

2) Ordinal Regression for sentiment analysis in twitterdata [20]

Approach

The proposed approach comprises of pre-handling tweets and utilizing a component extraction procedure that makes a conservative element This investigation expects to explain notion examination of twitter information concerning ordinal relapse utilizing many AI systems.inside the setting of this work, the paper blessing a methodology that means to separate Twitter estimation examination by building an evening out and evaluating model, a while later, ordering tweets into numerous ordinal classifications utilizing AI classifiers. Classifiers, similar to Multinomial stock relapse, Support vector relapse, call Trees, and Random Forest, are used right now. This methodology is advanced utilizing Twitter information set that is in open realistic inside the NLTK corpora assets. Test results show that Support Vector Regression ANd Random Forest have a for all intents and purposes comparative precision, that is more advantageous than that of the Multinomial inventory relapse classifier. Be that as it may, the decision Tree offers the absolute best precision at 91.81%. Trial results total that the anticipated model will find ordinal relapse in Twitter utilizing AI systems with a not too bad precision result. The exhibition of the model is estimated utilizing precision, Mean Absolute Error, and Mean square Error. inside the future, we endeavor to improve our methodology by attempting to utilize bigrams and trigrams. in addition, we will examine distinctive AI strategies and profound learning methods, similar to Deep Neural Networks, Convolutional Neural Networks, and repetitive Neural Networks[33]. Multinomial calculated relapse (SoftMax), Support Vector Regression (SVR), Decision Trees (DTs), and Random Forest (RF) calculations are utilized for opinion investigation characterization in the proposed system. For the genuine usage of this framework, a twitter dataset freely made accessible by the NLTK corpora assets is utilized. Exploratory discoveries uncover that the proposed approach can distinguish ordinal relapse utilizing AI strategies with great exactness. Also, results show that Decision Trees gets the best outcomes outflanking the various calculations.

Application

This methodology is best for Twitter assessment investigation Twitter slant examination is by and by a popular theme for investigation. Such examination is gainful on the grounds that it assembles and orders assessment by investigating gigantic social information. In any case, Twitter information have sure qualities that cause issue in directing assessment examination in differentiation to dissecting elective styles of information. Tweets are limited to one hundred forty characters, written in casual English, contain sporadic articulations, and contain numerous shortened forms and slang words to manage these issues, analysts have led considers that have some expertise in assessment examination of tweets.Smaller scale blogging sites have got one of the greatest web goals for people to communicate their considerations, sentiments, and perspectives with respect to totally extraordinary Topics.[34] Twitter is a generally utilized miniaturized scale blogging stage and interpersonal interaction administration that creates an enormous amount of information. Lately, analysts unmistakably made the utilization of social information for the assumption investigation of individuals' sentiments on an item, point, or occasion. Assumption examination, also called sentiment mining, is an essential normal language handling task. This strategy decides the assumption direction of a book as positive, negative, or unbiased.

3) Hybrid framework combining Lexicon based andmachine learning approach for sentiment analysis [21] Approach

The plan, advancement, and investigation of coordinated supposition examination system is clarified in detail.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Three totally various methodologies are used to feeling examination which fuses SWN, ML, and ml with GA streamlined component choice. they tend to proposed and built up a developmental model for highlight decision utilizing GA's transformative model. This tale approach brought about 36% - 42% diminished element size and about 5% duplicated strength when contrasted with a conventional ml approach. they tend to furthermore introduced a nearby examination of those methodologies with connection to totally unique datasets. in addition, our expounded investigation of different ml classifiers uncovered that the NB classifiers has the absolute best exactness (about 80%) though utilizing our GA put together generally ideal component decision with respect to Twitter and audits dataset while in the event of the geopolitical dataset, IB-k outflanked all the classifiers with the precision of 95%. in addition, they tend to assessed our proposed procedure for versatility by utilizing execution time correlation. they tend to establish that our framework indicated a straight speeding with the increased dataset size. In spite of the fact that, the time spent inside the decision of ideal list of capabilities abuse GA took in regards to 60% to 70% of the full execution time on audits dataset, hotheyver, it despite everything stayed straight and made a list of capabilities with 40% diminished size than the first element set.[35] GA based list of capabilities winds up in a speeding of displaying the classifiers up to 55% in order to show the upside of utilizing our component decrease rule over elective element decrease procedures; they've given Associate in Nursing precision correlation of GA based half and half methodology with PCA and LSA. The outcomes indicated that our GA based for the most part highlight decrease appeared at fifteen.4% expanded exactness over PCA and up to 40.2% expanded precision over LSA.

Application

This methodology has end up being a decent expansion inside the order of feeling mining. It gave the malleability of choosing among 3 wide utilized notion examination strategies as per custom wants. With additional preferences of GA based for the most part streamlining, it lessens include size and improves proficiency while keeping up the adaptability. later on, they will in general expect to expand this system for digital insight so it may encourage produce proposals for law-authorization organizations upheld client opinions.The Internet and web advancements have changed the manner in which our general public works . Interpersonal organizations, for example, Facebook and Twitter are currently ordinary for trading thoughts, sharing data, business, running political battles, and advancing items etc[36].. Web based life is commonly concentrated from alternate points of view i.e., gathering business knowledge for items advancement, checking malignant exercises for identifying andmoderating digital dangers, and supposition examination for dissecting individuals' input and audits.

4) Sentiment information in traditional TF-IDF algorithm and BiLSTM for improved word representation[22]

Approach :

Here, the commitment of the word's estimation measurements to printed content conclusion class is implanted into the customary TF-IDF calculation, and the weighted word vector is produced.

The development of the weighted word vector-Word2vec adaptation is utilized to accomplish apportioned portrayals of expressions. Word2vec innovation comprises of CBOW model and Skip-gram model. Both CBOW model and Skip-gram model comprise of enter layer, projection layer and yield layer. CBOW model predicts objective expressions fundamentally dependent on setting distribution.TF-IDF is a total of TF and IDF weight estimation techniques. It is the most by and large utilized weight estimation procedure in literary substance categorization..[37][38]The traditional neural network model is pointless in taking care of the grouping picking up information on in light of the fact that it's miles impractical to clarify the relationship between's the front and back of the succession. RNN (Recurrent Neural Networks) is a succession learning model that interfaces hubs among concealed layers and can contemplate grouping capacity dynamically.LSTM is straightaway, In the method of parameter preparing, the angle diminishes continuously till it vanishes. Accordingly, the span ofsuccessive information is restricted. Long Short Term Memory (LSTM) beats the issue of slope disappearance.At last, Sentiment examination of remarks based BILSTM is done so as to make every minute incorporate the setting realities, BiLSTM which consolidates bidirectional intermittent neural system (BiRNN) models and LSTM contraptions is utilized to hold onto the setting measurements. BiLSTM rendition treats all sources of info similarly. For the test of feeling assessment, the supposition extremity of the content to a great extent depends upon the words with conclusion data. The conclusion support of supposition word vectoris figured it out. Notion investigation commitments are essentially message order duties, and dispensed expression vectors don't consider the commitments of various words to the classification task. Right off the bat, the weighted word vectors are utilized in light of the fact that the contributions of BiLSTM adaptation, and the yields of BiLSTM model are utilized in view of the portrayals of the comment writings. At that point, the remark content vectors are contribution to the feedforward neural system classifier. At long last, the assessment inclination of the criticism is gotten.

Application:

Feeling investigation of comments particularly centers around the assumption direction examination of comment corpus, which shows that clients express positive, poor or unbiased suppositions towards product or occasions. Likewise, opinion investigation might be isolated intonews remark examination [2], item remark assessment [3], film comment assessment [4] and various sorts. These comments convey the points of view of Internet clients around stock, warm occasions, and so forth. Vendors can get a handle on the client fulfillment with the relevant item remarks. Potential clients can assess stock through survey these item remarks

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5) Integration of sentiment words and attention mechanism for improved sentiment classification [23]

Approach :

LSTM is an exceptional kind of RNN which can inquire about long haul conditions. RNN has a grouping of rehashing modules of simple structure. Be that as it may, the module of LSTM is increasingly perplexing. It doesn't just have one neural network layer however has four layers that connect in an exceptional manner. What's more, it has two states: concealed country and cell realm. The customary RNN form has deserts alongside evaporating inclination issue and detonating angle bother

. The disappearing slope will make the acing procedure extreme to unite, and detonating inclination will cause the shakiness of the examining procedure. Because of the deficiencies of the traditional RNN adaptation, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) has been proposed. Attention instrument is typically used in arrangement to grouping form to deal with encoder states. It likewise can be used in any assortment model to complete weight computations on past states. So as to beautify the commitment of basic words inside the literary substance representation, consideration instrument might be use[39][40][41]. By

utilizing the eye system, the outline of the enter succession isn't kept to the absolute last country. Supposition assessment on a colossal amount of writings underneath delicate measurements subjects can be performed to gain supervision and direction of popular conclusion. Here, Sensitive Information Topics-Based Sentiment Analysis Model (SITSAM) is proposed, which coordinates the point depiction of touchy records into the literary substance delineation picking up information on variant. Then, setting mindful intrigue system is conveyed to blast the commitment of assumption words to denounce conclusion arrangement.

Application:

The quick improvement of this subject is firmly identified with the enormous scope of web based life that show up on the web, for example, news surveys, conversation board conversations, online journals, Weibo, Twitter and interpersonal organizations. The data is significant for the two merchants and sellers since they regularly shouldknow about the appraisal and want of shoppers and the open around their product and administrations. For the administration, they also need to know the open's pointsof view and investigates on existing guidelines and up and coming rules. Through opportune input of those points of view, pertinent government choice producers can react rapidly. It is essential for popular assessment supervision to see touchy data themes by means of subject models and direct assumption examination fundamentally dependent on delicate realities topics[43]. In this manner, the material popular feeling following divisions can control the system general supposition and react to the necessities of netizens in an opportune way.

6) Hybrid framework for document level sentimentanalysis[24]

Approach :

MOCA comprises of four segments: semantic learning, mindful model, neural communitarian filtering ,and multi-target streamlining from base to top.The first objective is to extricate semantic implications from writings by utilizing two stages: (i)It changes over each expression into an installing vector in an inactive semantic territory and empowers to look at changed words, e.G., figuring their semantic likeness or separation on slant words, which benefits for the subsequent investigation contexts.(ii)A bidirectional repetitive neural system (BRNN) is followed to catch the semantic implications of a book which incorporates the word arrangement in both ahead and in reverse directions. Then a way to deal with compute the weighted aggregate of those states dependent on the consideration instrument is followed[44][45] . Specifically, this viewpoint misuses the consideration of client and thing to find significant expressions for them,in light of the fact that extraordinary clients have various options on phrases for communicating their sentiments, simultaneously as uncommon items furthermore have exceptional appropriate expressions for assessing them. Other than the literary substance inserting from the mindful model, the part furthermore thinks about individual other options and thing 's attributes, which

won't be reflected in the content and significantly influence the rating. This component devises another NCF system to adaptation the customer portrayal , thing portrayal , and their connections, seeing that clients show customized options on items.Multi-target enhancement (MOO) is utilized to display the record arrange conclusion investigation as every classification and relapse commitments in one unified structure, by method for limiting the classification and relapse misfortunes on the rating forecast simultaneously. The multi-objective streamlining licenses the 2 assignments to support each other.

Application:

The work utilizes the content measurements yet disregard the critical influence of the client and article records the client who composes the content and the item which is assessed inside the content. As a general rule, the purchaser and thing realities significantly impact the scoreof the buyer providing for the item. For instance, clients normally explicit the equivalent supposition with client specific words that depend upon their inclinations,propensities, and encounters; every thing has a lot of article specific contrasting words that rely upon the item's qualities.

7) Using Day-of-the-week effect, Sentiment classification and SVM for Stock Market prediction and Analysis.[25]

Approach

Stock market movement is a very dynamic and non-linear system where sentiment of investors constitutes a very crucial role. This sentiment is present in various news articles, blogs, interviews, social media websites and other text sources and can be extracted by the NLP concept of sentiment analysis. Also various companies offer data in form of stock market indices which is also helpful in predicting this movement.

Integrating the two and putting the aspects together in an SVM can yield accurate results which may be helpful in deciding when to buy or sell a particular stock. The stocks and their movements have earlier been forecasted manually or by use of mathematical models[46]. Later machine learning algorithms and models have also been used. Data of the stock market indices and trends are fed to this model in order to compute reliable predictions. However whatever investors think is the most important factor governing the market movements and this cannot just be calculated by some numerical data. This can be done through sentiment analysis of texts. Sentence based sentiment analysis can be used. The day of the week effect which says thatinvestor sentiment is low or negative on Mondays due to possible announcement of bad news on Fridays by companies, is also a factor that may affect the sentiment of an investor is also an important factor when computing the overall sentiment of an investor. Some previous researches also include the current affairs of the most common sports globally played so that it can become an important deciding factor for investor sentiment.Not only does this paper takes into account, the financial market data in order to make predictions, but sentiment analysis which incorporates investor psychology is also used. A sentiment index is prepared out of this data which is extracted from unstructured news resources. To increase reliability of the sentiment analysis, day-of-the-week effect, which is seldom included in the related works, is also taken into consideration. Since data in stock market is dynamic and non-linear, SVMs are used to make predictions and then adjustments are made to real market situations by use of rolling window approach. Predictions are then used to finalise important investment related decisions and to compare the market strategies.[47] The proposed solution can help reduce the risks for individuals and institutions. Due to these reduced risks, they can make more profitable investments in the market. The stop-loss order approach/method has also been employed in order to prevent any investment from going into loss beyond a certain non-recoverable point. Data from SSE 50 index is used. Every online blog and newsletter of the 51 shares from the Sina stock forum and Eastmoney stock forum over a timespan of 17th June, 2014 and 7th June, 2016, which includes exactly 485 days of open trading in the markets. Both of the platforms are highly known and popular discussion forums in China.

Application

As it is clearly evident from the given approach and research title itself, that this may be an extremely useful approach while analysing the trends in stock markets and the movement of existing stock prices. This analysis proves to be extremely useful to an investor while attempting to invest in a particular stock or betting on the movement of that stock[48]. A similar approach can be useful in other sectors of marketing and finance as well. For Example, another very dynamic yet extremely important factor in the market is the pricing of gold. Since there exist multiple purity categories of gold, and multiple factors concerning the gold production, mining and the economy that affect this pricing, an SVM integrating with other such governing factors may be useful to give an individual the supposedly precise analysis on the timeline of investments in gold that he should be making in gold which will fetch him the optimal profit. The sales of a particular product or service may also be extremely dynamic and dependent on the day-of-the-week and time of the year. Using SVM's, customer and distributor sentiments, the time where the product marketing needs a push, can be approximated. This will help minimizing loss for the company and better results. Oil prices globally depend on the OPEC and whatever goes on politically between the countries that are part of it. Using sentiment categorization of news articles from trustable sources, about OPEC and countries involved in it, it may be possible to determine the approximate rise or fall that may happen in oil prices. Since oil pricing affects transportation and process of other commodities aa well, predicting a trend in oil pricing may be extremely vital information.

8) Visitor Opinion Mining and Sentiment Classification [26]

Approach

Visitor surveys are data hotspots for explorers to think about vacationer places. Aspect based estimation grouping techniques have indicated guarantee in smothering the commotion. Be that as it may, very little research has been done on programmed perspective distinguishing proof, and ID of certain, rare and co- referential viewpoints, bringing about misclassifications. This model exhibits a system of aspect based assumption order that won't just recognize the aspects effectively yet can perform classification task with high precision. The travel industry is a developing industry with expanding significance to nations all inclusive. With the approach of cell phones, voyagers visit spots of intrigue and offer their conclusions on different social stages and sites. These suppositions present an expansive view to perusers about a vacationer place. Be that as it may, perusers might be confounded about whether to visit the visitor place because of decent variety of assessments. Sentiment analysis helps in broad classification of tourist sentiments as positive or negative. Every visitor place has various viewpoints and a basic twofold outcome is frequently deficient. Along these lines, aspect based assessment arrangement techniques can be helpful. There are three issues related with the principal undertaking of aspect recognizable proof: (1) distinguish the certain viewpoints,e.g., "Yesterday we went to XX café, each dish was sleek,wet and too salty." This announcement verifiably gives a sentiment about a significant perspective "nourishment" that was not referenced in the first content. (2) Identify the co-referential perspectives, which allude to a similar viewpoint that notices in conclusions with various word or equivalent words, for example environment and feel are co referential angles in light of the fact that bothallude to a similar viewpoint. (3) Identify the perspectives that happen rarely yet are regardless significant, for example drapes and bed are to some degree unordinary perspectives yet significant for inns.[49][50]

Different data sets are used for hotels and restaurants but the source of the data is unrevealed.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

According to the studies, data is taken from popular tourism websites. In Step 1 (Data Collection), vacationer surveys about visitor places like lodgings and eateries are gathered from various web based life stages and sites. Stage 2 (Data Pre-processing), noise suppression and excess, and cleaned audits are changed into sentences. Stage 3 (Aspect Identification) discovers perspectives from pre-processed datasets utilizing a half and half viewpoint recognizable proof technique. Stage 4 (Classification) utilizes machine learning to arrange the distinguished aspects into positive or negative sentiment. ROC graphs are plotted which represent the outcome of aspect-based sentiment classification. Bar charts/graphs are produced with theintention to contrast the algorithms/approaches in various parameters[51][52].

Application

This research provides a tourist or a visitor without any local relatives or friends, with trustworthy and genuine data of opinions regarding various hotels and restaurants. A visitor gets to explore places that have the cumulative positive statement of previously visited tourists.[66] So, the visitor doesn't have to spend time in finalising eateries and lodging. This also saves the traveller from the burden of going through a bad experience and enjoy the most of their vacation. Findings and its mobile application are not only of use to a tourist, but also extremely vital for business of these hotels and restaurant. Once there is availability of the cumulative sentiment of customers and visitors and this data is also available to those who run these eateries and hotels, they can take this as feedback and adapt to the customers' suite. For example, when the visitors appreciate a particular place for a particular service or product, even other business owners may try to incorporate that service or product in their respective businesses (restaurants or hotels). In case of these visitors not liking a given place for some reason, the particular business owner may then work to fix up things and work to avoid any negative feedback from the user. With increasing competitiveness and desire to satisfy the visitor, the visitor's options for a given product/service will increase and the quality standardswill also improve consequently.

9) TD-BERT and variations for emotion analysis [27]Approach

The problem is an experiment consisting of three target-dependent variations of BERT. There performances are evaluated and assessed in order to significantly prove how the TD-BERT algorithm can actually outperform the existing and widely used methods of sentimentclassification which include embedding based models. Targeted sentiment examination explores the sentiment polarities on given target specifies from input writings. Not the same as sentence level sentiment, it offers all the more fine-grained information on every element notice. Sentiment Analysis/Classification is a basic undertaking in Natural Language Processing (NLP).[53] It is critical for understanding client created message in news reports, item surveys, or social discussions. Aspect-Based Sentiment Analysis (ABSA) being a very precise procedure in notion examination, expects to recognize the sentiment (e.g., positive, negative, nonpartisan, conflict) of an aspect classification or an objective. It is important to recognize the aspect or a target in a sentence and also the sentiment/sentiments that are associated with the given targets. The size of computerized media is developing at a detonating speed, which makes data utilization a difficult undertaking. A huge part of the computerized media is client produced, yet physically finding the necessary data is past the capacity of any person. Machine helped media preparing is significant for some recipients, including governments, companies and people, while its applications incorporate stock value expectation, item suggestion, assessment of public sentiment, and so on. All these require exact extraction of primary entities, together with feelings or frames of mind communicated by the creator. BERT is a model/approach for representing language, that utilizes a bi-directional or 2-way Transformer system to train beforehand, a language model along with a huge corpus, and makes suitable changes in order to increase precision of the pre-prepared model on different assignments. TD-BERT takes output from the target terms (in red). In the case of a number of target words/phrases, a maximum pooling activity is taken before information given into the following completely associated layer. TD-BERT takes the situated outcome at the target terms as contribution for classification rather than the first tag. There are three datasets in the experiment. The first two are from Sem Eval-2014 task41 remembering information for the café and workstation spaces, which are generally utilized in past work. The lastone is an assortment of tweets gathered.

Application

Pre-prepared models have demonstrated to be of extraordinary assistance in achieving numerous NLP undertakings, for example, natural language induction, content grouping and question-answering systems. These ideal models contain a semi-supervised language model prepared on huge corpora of information; they are later tweaked to deal with downstream errands. Since its first appearance, BERT has increased a great deal of ubiquity in the scholarly network. It has been applied to different NLP errands, including content classification for question answer system[54]. The first work by Devlin et al. (2018) contained outcomes on BERT's working efficiency over the Stanford Question Answering Dataset task (Rajpurkar et al., 2016), where the framework needed to foresee the appropriate response length for a particular question in a Wikipedia section. Yang et al. (2019) went further, making a question answer framework that was conveyed as a chatbot. In any case, both these investigations handled the errand of open space question replying, while we center around situations where BERT was misused to create frameworks for genuine applications. For example, Lee et al. (2019) made another BERT language model pre-prepared on the biomedical field to tackle space explicit content mining undertakings (BioBERT). Its outcomes are amazing, yet BioBERT is skilled to performwell on space explicit information in view of its enormous pre-preparing process. While the pre-preparing clearly yields better exhibitions, it is exceptionally costly with respect to computational expenses and time utilization. Both SciBERT and BioBERT likewise presentarea explicit information for pre-preparing. Beltag et al. utilize 1.14M researches are irregular pick from SemanticScholar to tweak BERT and building Sci BERT.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The corpus incorporates 18% software engineering space study and 82% wide biomedical area researches. Then again, Lee et al. utilize BERT's unique preparing information which incorporates English Wikipedia and BooksCorpus and space explicit information which are PubMed modified works and PMC full content articles to calibrating BioBERT model.

10) Emotion recognition of tweets specific to football.

Approach

The issue is to distinguish the assessment or feeling behind football explicit tweets we build up a football- explicit supposition dataset which we name physically. We at that point use our dataset to consequently make a football-explicit slant lexicon[55][56]. At last, we build up a conclusion classifier which is equipped for
perceiving opinions communicated in football discussion. We lead broad analyses on our dataset to think about the exhibition of various learning calculations in recognizing the notion communicated in football related tweets.
Sentiment classification is then done by:
BAG-OF-WORDS (BOW) is one of the most popular representations of textual data and is widely used in text classifcation. Given apredened set of vocabulary, generated using a word or a sequence of words, a document is represented as an N-dimensional feature vector. Each element *xi* in the feature vector corresponds to a word *wi* in the vocabulary. The value of *xi* can be a binary value that indicates the appearance of a word in the document *di* or the number of occurrence of *wi* in *di* that indicates its term frequency (TF). PART-OF-SPEECH (POS) features are commonly used in sentiment analysis. POS taggers determine the part of speech of each word in a sentence and labels it as noun, verb, adjective. etc. Various EXISTING SENTIMENT LEXICONS have been developed using diverse resources; however, the limited coverage of sentiment words in a single lexicon is one of the major limitations. Moreover, large numbers of existing lexicons do not contain the abbreviations, emoticons, and slang widely used in social media. Then there are several text classifiers used like Support vector machine classifier (SVM), Multinomial naïve bayes classifier, random forest.

Application

Sports fans produce a lot of tweets which mirror their assessments and sentiments about what's going on during different games. Given the fame of sports occasions, in this work, we centre around examining slant communicated by sports fans through Twitter. These tweets mirror the adjustments in the fans' feeling as they watch the game and respond to the occasions of the game,e.g., objective scoring, penalties, wickets, runs etc. There are many other applications while we can mine the sentiments of users and find patterns using machine learning or artificial intelligence and predict events in a particular game. Not only football but sentiment analysis can be done for any game by using different datasets and then predictions about events can be made and then trust mining of users can be done to make online friends with people who have similar opinions or sentiments.[57]

11) Finding user trust patterns on e commerce platforms to betterly review products.

Approach

The comments of users about a product are analysed on an e-commerce platform and people with similar opinions on the product form a bigger opinion as customers with the same opinion tend to trust them more rather than those whose opinions differ. Following this perspective on point, an E-commerce framework surveys mining focused assessment comparability investigation approach is advanced to investigating clients' closeness and their trust. We partition the trust into two classes, to be specific direct trust, and spread of trust, which speaks to a trust indication between two people. The immediate trust degree is acquired from conclusion and closeness, and we present a substance estimation word pair digging strategy for likeness highlight extraction. The engendering of trust is determined by the transitivity include. Utilizing the proposed trust portrayal model, we utilize the shortest path algorithm by the djikstras algorithm way to depict the tightness or closeness of trust and set forward an improved better and error free way calculation to make sense of the spread trust between clients and buyers. A huge scale E-commerce site surveys dataset is gathered to look at the precision of the calculations and possibility of the models. The exploratory outcomes show that the supposition comparability investigation can be aproficient technique to discover trust between clients in E-commerce platforms. for direct trust calculation, that is, coordinated trust in the work, we utilize the weighted normal technique to register them, which is like other existing works. Be that as it may, simultaneously, we present a going with factor of feeling, the rating which generally exists in E-business audits, for loads assessment. Which is, the immediate trust figuring affected by the realities whether the clients have the equivalent wistful inclination or not for something very alike.[58] for non direct trust calculation, which is balanced trust however a third ones, we present chart based proliferation calculation. in light of the proposed trust portrayal model, we utilize a most brief way to depict the snugness of trust and set forward an improved concise way calculation to make sense of not direct trust connection between clients. The proliferation trust is processed by coordinating the adjecent trust dependent on briefest way processing and computation.

Application

Mining trust of individuals can be advantageous for some applications. Film suggestions can be enhanced the premise of remarks it gets. Clients can post remarks about motion pictures and afterward every film can be doled out a rating based on conclusions of individuals. Thiscalculation can be utilized on dating site stages where individuals can be mapped based on contributions about their writings and interests to assist them with discovering individuals of comparative taste and trust.[59]

12) Social media sentiment analysis and application inautomobile sales prediction and business

Approach

Internet based life is blasting, offering remarks or communicating suppositions about Online products are simpler than any time in recent memory. Information from web based life can be one of the important contributions To gauge vehicle deals.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Moreover, there are some different impacts, for example, securities exchange esteems On the buying intensity of vehicles. Two multivariate relapse models with web based life information And securities exchange esteems and time arrangement models are utilized to appraise month to month complete vehicle deals. The Least squares vector relapse (LSSVR) models are utilized to manage multivariate relapse information. Three sorts of information are utilized, to be specific tweets slant scores, securities exchange esteems and cross breed information to assess month to m o n t h absolute vehicle deals in the USA. Contains two nostalgic scores of half breed information Tweets and financial exchange esteems. Furthermore, regular components of month to month complete vehicle deals are utilized Decentralizing both month to month absolute vehicle deals and each of the three kinds of info information. There are time arrangement models[60] Naive model, exponential smoothing model, autoregressive coordinated moving normal model, Periodic Autoregressive Integrated Moving Average Model, and Backpropagation Neural Networks and LSSVR with time arrangement models.[61] The numerical outcomes recommend the utilization of half and half information with deconvolution LSSVR models can deliver more exact outcomes than different models with various information. In this way, both internet based life information and stock qualities are required to gauge month to month absolute vehicle deals; And Decentralizing systems can improve figure exactness in evaluating month to month complete vehicle sales.[62][63]

Application

This algorithm can be used for many prediction purposes like prediction of stock market values of shares and commodities. T h i s can be done by different datasets of commodities. then the dataset of OPEC or oil and natural gas corporation can be used to predict the sales of crude oil, petrol, diesel, coal, etc . this can also be done by mining and finding patterns in investor , customer, and retailer sentiments. [64][65]

III Evaluation methods sectionCommon evaluation methods-

1) Accuracy -Accuracy is s ratio of correctly predicted observation to the total observations.Accuracy = TP+TN/TP+FP+FN+TN

2) Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Precision = TP/TP+FP

3) Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to all observations in actual class - yes.

Recall = TP/TP+FN

4)F1-Score

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

5) Mean Absolute Error -MAE measures the average magnitude of the errors in a set of predictions, without considering their direction.

6) Mean Squared Error - Root mean squared error (RMSE): RMSE is a quadratic scoring rule that also measures the average magnitude of the error. It's the square root of the average of squared differences between prediction and actual observation.

7) TF-IDF

TF − IDF (t, d) = TF (t, d).IDF (t)

TF_IDF is u s e d to evaluate the significance of a word to a document in a dataset. Each word is assigned a weight in the document

Datasets -

1. **ChnSentiCorp-Htl-ba-10000** This dataset includes 10000 piece of hotel reviews, labeled as positive and negative.
2. **NLPCC-ECGC** dataset come from online comments on Weibo, which includes 1 million dialogues in the post-response form.
3. **UCL ML**- Repository's Sentiment Analysis Dataset with IMDB reviews.
4. **Twitter Dataset** – Tweets collected by twitter API.
5. **Geopolitical Dataset** - Dataset related to 2016 US presidential election

V. RESEARCH METHODS

Table-I Comparison of the various sentimentanalysis methods based on various parameters

S. No	Ref.	Methodology Domain	Problem Addressed	Evaluation Parameters	Datasets
1	Sentiment analysis of short text data with a bi level attentionmodel [19]	Machine Learning with bi- level attention mechanism	Difficulties in sentiment analysis due to data sparsity Topic modelling of short text data	Accuracy F1 Score MCNN Sequence attention	Chn Senti Corp-Htl-ba-10000 NLPCC-ECGC
2	Ordinal Regression for sentiment analysis in twitter data [20]	Sentiment analysis using ordinal Regression	Ordinal classification, opinion mining in twitter Sentiment orientation	PrecisioRecallF1-Score Support Accuracy Mean Absolute Error Mean Squared Error	NLTK Twitter Dataset(using API)
3	Hybrid framework combining Lexicon based and machine learning approach for sentiment analysis [21]	Hybrid framework using genetic algorithm for sentiment analysis	Lack of labeled data in lexicon based approaches for sentiment analysis,	Accuracy Precision Recall	UCLML Repository's, Dataset with IMDB reviews Twitter Dataset, Geopolitical Dataset related to 2016 US presidential

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

			Less Accuracy in machine learning models for sentiment analysis	F-measure	election
4	Sentiment information in traditional TF-IDF algorithm and BiLSTM for improved word representation [22]	A conclusion investigation technique for remarks in light of BiLSTM is proposed	Improved word representation method is proposed to compute the sentiment information ofword	Precison Recall F-measure	15000 hotel comment texts (Data set) crawled from Ctrip 300-dimensional word vectors trained by the Skip-Gram model provided by Data Science .
5	Combination of sentiment words and attention mechanism for improved sentiment classification [23]	Attention mechanism with neural network, and context aware vector are used to calculate the weight of each word.	Mining the slant propensities contained in enormous messages on the Internet.	Accuracy Loss rate%	News reports on the "Shipwreck in Phuket,Thailand. Hotel review data set provided by Tan Songbo
6	Hybrid framework for document level sentiment analysis[24]	Multi Objective, Collaborative and Attentive framework calledMOCA	Document-level sentiment analysis to predict the scores of ratings given by users of anyitems/products.	Accuracy Root Mean Square Error	IMDB and Yelp Dataset Challenge in2013 and 2014
7	Using Day-of-the- week effect, Sentiment classification and SVM for Stock Market prediction and Analysis[25]	Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine	Forecasting stock market trends using Sentiment analysis	Accuracy	Data from SSE 50index
8	Visitor Opinion Mining and Sentiment Classification[26]	Aspect-Based sentiment Analysis	The tourists' and travelers' reviews or sentiments on various online platforms is used to determine the cumulative sentiment for a particular touristplace	Accuracy Precision Recall F-Measure	Data from popular tourism websites
9	TD-BERT an d variations for emotion analysis[27]	Target Dependent Bidirectional Encoder Representations from Transformers	Sentiment analysis or classification using BERT and its Target-Dependent versions.	Accuracy F-measure	Sem Eval-2014 task41 Assortment of tweets
10	Emotion recognition of tweets specific to football. [28]	Football-specific sentiment lexicon	To identify the sentiment or emotion behind football specific tweets	Accuracy F-measure	FIFA World Cup(FIFA) 2014 Champions League (CL) 2016/2017
11	Finding user trust patterns on e commerce platforms to betterly review products. [29]	Entity-sentiment word pair mining method for similarity featureextraction	E-commerce system reviews mining oriented sentiment similarity analysis approach is put forward to exploring users' similarity andtheir trust.	Accuracy Precision Recall F-measure	Dataset collected fromAmazon.
12	Social media sentiment analysis and application in automobilesales prediction andbusiness [30]	Least squares support vectorregression (LSSVR)	Multivariate regression models with social media data and stock market values and time series models are employed to predict monthly total vehicle sales.	Actual and predicted values, Accuracy	Twitter Data (from API) Data from SSE 50 index

Table 1 gives a comparison of the different approaches for sentiment analysis and the evaluation parameters the different approaches have been evaluated upon and compared to pre-existing research methods. Methods in [20] ,[21] and [30] proposed approaches tested and trained on datasets extracted from twitter. Earlier research methods of sentimental analysis had ordinal classification. Sentiment analysis using ordinalregression introduced in study[20] addressed this problem and was comparatively better in parameters like precision recall , accuracy and F1 score. Hybrid framework using genetic algorithm for sentiment analysis introduced in model [21] was tested and trained on US presidential election 2016 dataset and addressed the issue of lack of labelled data in lexicon based approaches for sentiment analysis, less Accuracy in machine learning models for sentiment analysis. One research[3] introduced an approach to improve multivariate regression models with social media data and stock market values and time series models are employed to predict monthly total vehiclesales. After the use of LSSVR model , actual and the predicted values accuracy came much closer than before. In the era of big data , mining massive texts from the internet is much difficult and there was a huge loss percentage during the extraction of the data until attention mechanism with neural network, and context-aware vector is acquainted with ascertain the heaviness of each word.in model [23] was tried on News covers the "Wreck in Phuket, Thailand.Hotel audit informational collection gave by Tan Songbo. A slant investigation technique for remarks dependent on BiLSTM is proposed in strategy [22] dependent on 15000inn remark writings (Data set) slithered from Ctrip 300- dimensional word vectors prepared by the Skip-Gram model gave by Data Science. Improved word representation had better scores at precision recall and F1 scores among its peer research approaches as it had better context and explicit information capturing.

Models in [24] and [21] were compared on the basis of accuracy and root mean , both models used IMDB datasets to review public rating on various series and films. In mehtod[24] , a study on new hybrid framework MultiObjective, Collaborative and Attentive framework called MOCA was done for document-level sentiment analysis to predict the scores of ratings given by users of any items/products. Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine was studied in study of [25] as the finance market depends a lot on public trends and emotions and is swayed by it. To accurately predict the future market shares of the companies. The study helped in increasing the accuracy of prediction models exponentially. Aspect-Based sentiment Analysis introduced in method [26] to analyse tourists' and travelers' reviews or sentiments on various online platforms is used to determine the cumulative sentiment for a particular tourist place. Data was abstracted from various websites as datasets. There was even field specific study done like in study[28] to identify the sentiment or emotion behind football specific tweets ,football-specific sentiment lexicon was created. Field specific studies help in increasing F-scores.

Table – II Pros/Cons of base methods of sentiment analysis

S. No.	Ref.	Advantages/Pros	Limitations/Cons
1	Sentiment analysis of short text data with a bi level attention model [19]	The performance of the proposed model in the case shows that BAM is effective in short text sentiment analysis.	One major error type come from the failure to understand the priority of local features. The model pay too much attention on the positive expression
2	Ordinal Regression for sentiment analysis in twitter data [20]	This approach tackles sentiment analysis on tweets full of noise. Another advantage of this model is that technique for sentiment polarity is also different.	The disadvantage of the proposed solution is that it does not take into consideration the computational and time complexity into consideration.
3	Hybrid framework combining Lexicon based and machine learning approach for sentiment analysis [21]	This framework for sentiment analysis has proved to be a high- quality addition in the discipline of opinion mining. It provided the flexibility of choosing among 3 broadly used sentiment analysis strategies in step with custom needs	A disadvantage of this technique is that it integrates two methods which increase the levels and time and space complexity and computational complexity
4	Sentiment information in traditional TF-IDF algorithm and BiLSTM for improved word representation [22]	BiLSTM model completely thinks about the setting data and can more readily get the content portrayal of the remarks. At last, through the feedforward neural system and softmax mapping, the supposition propensity of the content is gotten	The sentiment analysis method of comments based on BiLSTM consumes a long time in the training model
5	Combination of sentiment words and attention mechanism for improved sentiment classification [23]	Test results affirm that the system of feeling word reference labeling is plausible. blend of conclusion expressions and consideration instrument has a huge enhancement for the feeling grouping.	For the case of a small quantity of texts, this method can't get enough texts as the training statistics
6	Hybrid framework for document level sentiment analysis[24]	MOCA has best execution as far as precision. It is viable in demonstrating the client and thing data with multi-target streamlining, MOCA accomplishes the best outcomes on the three datasets	As λ for the classification task increases (i.e., the loss weight for the regression task decreases), both accuracy and MOCA becomes worse .
7	Using Day-of-the-week effect, Sentiment classification and SVM for Stock Market prediction and Analysis[25]	The proposed solution can help reduce the risks for individuals and institutions. Due to these reduced risks, they can make more profitable investments in the market.	The proposed approach fails to detect the dramatic decline at first, but predicts the trend afterward, and is able to uncover the following rise.
8	Visitor Opinion Mining and Sentiment Classification.[26]	A tourist will save their time by not visiting bulk of websites and not reading all reviews to understand the sentiment of the travelers that have already been to the destination.	The application's efficiency in terms of battery consumption can be an issue.
9	TD-BERT and variations for emotion analysis [27]	As far as classification exactness, TD- BERT is 3-11% higher, particularly on the workstation dataset	There are yet certain classes of data, for which the present model can't give an acceptable arrangement.
10	Emotion recognition of tweets specific to football. [28]	The algorithms or tools used are very specific and accurate among the rest and provide us with 72% accuracy	Algorithm is very hard to understand for novice people and that not normal texts but only tweets can be analyzed.
11	Finding user trust patterns on e commerce platforms to betterly review products. [29]	The data set is very diverse and has the following attributes so each user can find another user accurately with a common trust pattern	Not each user gives their reviews on each item, so the user's reviews data are usually sparse for a particular item. The degree to which people trust others is different for different things.
12	Social media sentiment analysis and application in automobile sales prediction and business [30]	The numerical results indicate that using hybrid data with DE seasonalizing procedures by the LSSVR models can obtain more accurate results than other models with different data.	The determination of keywords for Twitter significantly affects the search results of tweets and have influences on forecasting accuracy

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

VI. CONCLUSION AND FUTURE DIRECTIONS

This research makes a survey of algorithms and enhancements in sentiment analysis. The reasearch also discusses the various applications to which the discussed approaches can be applied to. The anlysis also discusses the various evaluation methods and the datasets used. [19] Presents the calculation (Bi-level consideration system) which can be best utilized in short messages One significant mistake type originate from the inability to comprehend the need of nearby highlights. The model give a lot of consideration on the positive expression.[20] Tackles opinion examination on tweets brimming with clamor. Another bit of leeway of this model is that method for assumption extremity is additionally unique. The disservice of the proposed arrangement is that it doesn't think about the computational and time intricacy into consideration.[21] end up being a top notch expansion in the order of sentiment mining. It gave the adaptability of picking among 3 extensively utilized opinion investigation procedures in step with custom needs yet it incorporates two strategies which increment the levels and existence intricacy and computational unpredictability .[22] model completely thinks about the setting data and can all the more likely acquire the content portrayal of the remarks. At last, through the feedforward neural system and softmax mapping, the slant propensity of the content is gotten however expends quite a while in preparing model.[23] Confirm that the method of assumption word reference labeling is achievable. blend of slant expressions and consideration system has a critical enhancement for the feeling grouping. For the instance of a little amount of writings, this technique can't get enough messages as the preparation statistics.[24] has best execution as far as exactness. It is viable in demonstrating the client andthing data with multi-target streamlining yet As λ for the characterization task increments (i.e., the lossweight for the relapse task diminishes), both precision and MOCA turns out to be more regrettable. In [25] The proposed arrangement can help decrease the dangers for people and establishments. Because of these diminished dangers, they can make increasingly productive interests in the market however The proposed approach neglects to identify the emotional decay at first, yet predicts the pattern a short time later, and can reveal the accompanying ascent.

In [26] A visitor will spare their time by not visiting greater part of sites and not perusing all audits to comprehend the supposition of the explorers that have just been to the goal. The application's productivity as far as battery utilization can be an issue. To the extent classification precision, TD-BERT is 3-11% higher, especially on the workstation dataset.[27] however There are yet sure classes of information, for which the current model can't give a worthy course of action. In [28] The calculations or devices utilized are quite certain and exact among the rest and give us 72% exactness yet Algorithm is difficult to comprehend for beginner individuals and that not ordinary messages but rather no one but tweets can be broke down. In [29] The informational collection is different and has the accompanying qualities so every client can discover another client precisely with a typical trust design Small scale blogging sites have got one of the greatest web goals for people to communicate their musings, sentiments, and mentalities with respect to totally unique Topics[23]. Twitter is a generally utilized small scale blogging stage and informal communication administration that produces a monstrous amount of information. Lately, specialists undeniably made the utilization of social information for the assessment examination of individuals' feelings on an item, subject,or occasion. Conclusion investigation, moreover called supposition mining, is a pivotal common languagehandling task. This strategy decides the assessment direction of a book as positive, negative, or unbiased. Estimation examination has interesting applications and has levels of popularity as of late. beside associations for building up the product , clients likewise are a lot of needing to recognize the audits from a few terms upheld what the corporate needs to attempt to and consequently the choices to sell the product by contrasting the assessments of a few organizations[24]. Feeling investigation will get the criticism bolstered the situation patterns looked by the fluctuated fields. Up and coming ages need a more profound examination and thinking for restricting a few proficient perspectives bolstered human considerations and brain science which overcomes any issues among unstructured and unimportant data[46].

Accordingly we've considered different methodologies of the numerous creators sees which give a few difficulties that emerge to the sheer measure of information inside the web and it demonstrates to call attention to that the slant investigation might be a creating zone

REFERENCES

1. Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177).
2. Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27, 16-32.
3. D. D. Droba, “Methods used for measuring public opinion,” American Journal of Sociology, pp. 410–423, 1931.
4. “Public Opinion Quarterly.” [Online]. Available:
5. //poq.oxfordjournals.org. [Accessed: 02-Dec-2016].
6. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® inInformation Retrieval, 2(1–2), 1-135.
7. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
8. Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture (pp. 70-77).
9. Gautam, G., & Yadav, D. (2014, August). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh International Conference on Contemporary Computing (IC3) (pp. 437-442). IEEE.
10. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
11. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.
12. O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., & Smeaton, A. F. (2009, November). Topic-dependent sentiment analysis of financial blogs. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion (pp. 9-16).
13. Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012, August). Sentiment analysis on social media. In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 919-926). IEEE.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

14. Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.
15. Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.
16. Kang, H., Yoo, S. J., & Han, D. (2012). Senti- lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Systems with Applications, 39(5), 6000-6010.
17. Cakra, Y. E., & Trisedya, B. D. (2015, October). Stock price prediction using linear regression based on sentiment analysis. In 2015 international conference on advanced computer science and information systems (ICACSIS) (pp. 147-154). IEEE.
18. Mullen, T., & Collier, N. (2004, July). Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the 2004 conference on empirical methods in natural language processing (pp. 412-418).
19. Dos Santos, C., & Gatti, M. (2014, August). Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 69-78).
20. Liu, W., Cao, G., & Yin, J. (2019). Bi-Level Attention Model for Sentiment Analysis of Short Texts. *IEEE Access*, *7*, 119813-119822.
21. Saad, S. E., & Yang, J. (2019). Twitter Sentiment Analysis Based on Ordinal Regression. *IEEE Access*, *7*, 163677-163685.
22. [21]Iqbal, F., Hashmi, J. M., Fung, B. C., Batool, R.,
23. Khattak, A. M., Aleem, S., & Hung, P. C. (2019). A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction. *IEEE Access*, *7*, 14637-14652.
24. [22] Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment Analysis of Comment Texts Based on BiLSTM. IEEE Access, 7, 51522-51532.
25. [23]Xu, G., Yu, Z., Chen, Z., Qiu, X., & Yao, H. (2019).
26. Sensitive Information Topics-Based Sentiment Analysis Method for Big Data. IEEE Access, 7, 96177-96190.
27. Zhang, J. D., & Chow, C. Y. (2019). MOCA: Multi- objective, collaborative, and attentive sentiment analysis. IEEE Access, 7, 10927-10936.
28. Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Systems Journal, 13(1), 760-770.
29. Afzaal, M., Usman, M., & Fong, A. (2019). Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews. IEEE Transactions on Consumer Electronics, 65(2), 233-242.
30. [27] Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-Dependent Sentiment Classification With BERT. IEEE Access, 7, 154290-154299.
31. Aloufi, S., & El Saddik, A. (2018). Sentiment identification in football-specific tweets. *IEEE Access*, *6*, 78609-78621.
32. Zhang, S., & Zhong, H. (2019). Mining users trust from e-commerce reviews based on sentiment similarity analysis. *IEEE Access*, *7*, 13523-13535.
33. Pai, P. F., & Liu, C. H. (2018). Predicting vehicle sales by sentiment analysis of Twitter data and stock market values. *IEEE Access*, *6*, 57655-57662.
34. Yang, K., Cai, Y., Huang, D., Li, J., Zhou, Z., & Lei,
35. X. (2017, February). An effective hybrid model for opinion mining and sentiment analysis. In 2017 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 465-466). IEEE.
36. Fu, X., Yang, J., Li, J., Fang, M., & Wang, H. (2018). Lexicon-enhanced LSTM with attention forgeneral sentiment analysis. IEEE Access, 6, 71884- 71891.
37. Jiang, F., Liu, Y. Q., Luan, H. B., Sun, J. S., Zhu, X., Zhang, M., & Ma, S. P. (2015). Microblog sentimentanalysis with emoticon space model. Journal of Computer Science and Technology, 30(5), 1120-1129.
38. Yang, Q., Rao, Y., Xie, H., Wang, J., Wang, F. L., Chan, W. H., & Cambria, E. C. (2019). Segment-level joint topic-sentiment model for online review analysis. IEEE Intelligent Systems, 34(1), 43-50.
39. [35] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.
40. Jain, A. P., & Dandannavar, P. (2016, July). Application of machine learning techniques to sentiment analysis. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 628-632). IEEE.
41. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.
42. Bouazizi, M., & Ohtsuki, T. (2017). A pattern-based approach for multi-class sentiment analysis in Twitter. IEEE Access, 5, 20617-20639.
43. Gutierrez, P. A., Perez-Ortiz, M., Sanchez- Monedero, J., Fernandez-Navarro, F., & Hervas- Martinez, C. (2015). Ordinal regression methods: survey and experimental study. IEEE Transactions on Knowledge and Data Engineering, 28(1), 127-146.
44. Niu, Z., Zhou, M., Wang, L., Gao, X., & Hua, G. (2016). Ordinal regression with multiple output cnn for age estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4920-4928).
45. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.
46. Khan, A., Baharudin, B., & Khan, K. (2011). Sentiment Classification Using Sentence-level Lexical Based. Trends in Applied Sciences Research, 6(10), 1141-1157.
47. Fornacciari, P., Mordonini, M., & Tomaiuolo, M. (2015). A case-study for sentiment analysis on twitter. In WOA (pp. 53-58).
48. Kouloumpis, E., Wilson, T., & Moore, J. (2011, July). Twitter sentiment analysis: The good the bad and the omg!. In Fifth International AAAI conference on weblogs and social media.
49. Govindarajan, M. (2013). Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. International Journal of Advanced Computer Research, 3(4), 139.
50. J.Li, "Chinese sentiment dictionary," Research Institute ofInformation Technology, Tsinghua University, No.30 ShuangqingRoad,HaidianDistrict,Beijing,China
51. Irsoy, O., & Cardie, C. (2014, October). Opinion mining with deep recurrent neural networks. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 720-728).
52. S.Wang et al. , "Sentiment classfication of Uyghur text based onBLSTM", Computer Engineering and Design, vol.38, no.10,pp.2879-2886, 2017
53. Liu, N. (2013). The Research of Microblogging Short Text Oriented Sentiment Analysis [D]. Wu HanUniversity.
54. Dos Santos, C., & Gatti, M. (2014, August). Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 69-78).
55. L.Jiang, M.Yu, M.Zhou, et al., "Target-dependent Twitter sentiment classification," in Meeting of the Association for Computational Linguistics: Human Language Technologies," in Association for Computational Linguistics, Portland, OR, USA, Jun.2011, pp.151-160.
56. O.Irsoy, and C.Cardie, "Opinion Mining with Deep Recurrent Neural Networks," in Conference on Empirical Methods in Natural
57. R.Socher, A.Perelygin, J,Y.Wu, et al., " Recursive deep models forsemantic compositionality over a sentiment treebank," in Conferenceon empirical methods in natural language processing, Seattle, United States, Oct.2013, pp.1631-1642
58. Zhang, J. D., Chow, C. Y., & Zheng, Y. (2015, October). ORec: An opinion-based point-of-interest recommendation framework. In Proceedings of the 24th ACM International on Conference on Information andKnowledge Management (pp. 1641-1650). ACM.
59. Ganu, G., Kakodkar, Y., & Marian, A. (2013). Improving the quality of predictions using textual information in online user reviews. Information Systems, 38(1), 1-15.
60. Chen, T., Xu, R., He, Y., Xia, Y., & Wang, X.(2016). Learning user and product distributed representations using a sequence model for sentiment analysis. IEEE Computational Intelligence Magazine, 11(3), 34-44.
61. D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in ACL, 2015
62. Chen, T., Xu, R., He, Y., Xia, Y., & Wang, X.(2016). Learning user and product distributed representations using a sequence model for sentiment analysis. IEEE Computational Intelligence Magazine, 11(3), 34-44.
63. Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. Journal of economic perspectives, 21(2), 129-152.

64. Berument, M. H., Ceylan, N. B., & Ogut-Eker, G. (2009). Soccer, stock returns and fanaticism: Evidence from Turkey. The social science journal, 46(3), 594-600.
65. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of finance, 62(3), 1139-1168.
66. Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Stanford University, CS229(2011)http://cs229. Stanford. edu/proj2011/Goel Mittal- Stock Market Prediction Using Twitter Sentiment Analysis. pdf), 15.
67. Gillam, R. A., Guerard Jr, J. B., & Cahan, R. (2015). News volume information: Beyond earnings forecasting in a global stock selection model. International Journal of Forecasting, 31(2), 575-581.
68. Marrese-Taylor, E., Velásquez, J. D., Bravo- Marquez, F., & Matsuo, Y. (2013). Identifying customer preferences about tourism products using an aspect-based opinion mining approach. Procedia Computer Science,22, 182-191.
69. Zhu, J., Wang, H., Tsou, B. K., & Zhu, M. (2009, November). Multi-aspect opinion polling from textual reviews. In Proceedings of the 18th ACM conference on Information and knowledge management (pp. 1799- 1802). ACM.

AUTHORS PROFILE

Aayush Gupta, is a 3rd year Computer Science student at VIT, Vellore, India. His research interests are natural language processing, Big data. He is also interested in Cloud computing.

Anant Gandhi, is a 3rd year Computer Science student at VIT, Vellore, India. His area of interests include artificial intelligence and machine learning.

Saarthak Agarwal, is a 2nd year Computer Science Student with specialisation in bioinformatics at VITUniversity Vellore, India.

Shamin Chokshi, is a 2nd year Computer Science Student with specialization in bioinformatics at VIT University Vellore, India.

Saravanakuamr K is a professor at VIT, Vellore India. His area of research includes Information retrieval, Question answering System and Natural Language Processing