



ISSN: 2321-2152

**IJMECE**

*International Journal of modern  
electronics and communication engineering*

E-Mail

[editor.ijmece@gmail.com](mailto:editor.ijmece@gmail.com)

[editor@ijmece.com](mailto:editor@ijmece.com)

[www.ijmece.com](http://www.ijmece.com)

# Machine Learning-Based Sentiment Analysis of Food Reviews

<sup>1</sup> Sathuluri Beulah, <sup>2</sup> T. Bhagyasri

<sup>1</sup>Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.

<sup>2</sup> MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

## Abstract –

The use of the internet for interpersonal communication has become the norm. In casual conversation, it's implied that you utilize web-based life apps and locations. One of the most popular forms of online media in use today is Twitter. A lot of the things we do on a daily basis involve people sharing their preferences with a post. It is now usual practice to use supposition analysis. However, there are still a number of problems with stable Twitter thinking depiction execution, such as the following: illustrated extravagance challenges for emotional signals, generous class lopsidedness in a multi-class problem, and the use of various ordinary semantic models. Because many forms of online life evaluation depend on precise concealed Twitter ideas, these problems are dangerous. For Twitter idea study, a book examination structure is suggested, as is apparently necessary. This recording popularizes the use of Twitter data for estimation purposes. The words and expressions used in electronic systems administration media convey the viewpoints of individuals about various objects, companies, governments, and occurrences. The goal of the natural language processing (NLP) area's suspicion evaluation job is to remove favorable, unfavorable, or neutral polarity from names of electronic life material. Experts are compelled to complete their examination in assumption examination due to the remarkable expansion in solicitations for business affiliations and governments. The creation of product review analysis is approached in this investigation via the use of three state-of-the-art ML classifiers: SVM, Logistic Regression, Random Forest, and Naive Bayes. The datasets used for the testing are those collected from Twitter's Yelp. This information may be found on the internet. Their general attitude or preference towards the topic is reflected in the research or published papers, which are part of the extensive resource assessment that includes the debate, review objections, and destinations.

Index Terms – Machine learning, Food review, NLP, Sentiment Analysis.

## INTRODUCTION

Thoughts permeate every aspect of our life and have the power to fundamentally alter the way we do things. Everyone must be aware that there comes a time in any evaluation when a decision must be made. In this world, connections are essential for understanding how buyers typically feel about products and companies. Explicit buyers are usually expected to be aware of the general public's opinions on a product before purchasing it, just as voters are expected to be aware of the general public's preferences on political candidates before casting a ballot in an election. Friends and family are the primary sources of evaluations for people at the moment. Assessment reviews and focus groups are immediately considered by coalitions and businesses when they need consumer hypotheses. It has long been a massive enterprise for advancement, development, and political purpose connections to guarantee out buyer and open assumptions. In spite of the fact that electronic structures association is in a precarious state, people and organizations are still using the content as if it were a touch of these media in places like studies, gathering exchanges, regions, Twitter, comments, and social alliance sites on the web. If you need to buy something unique these days, you don't have to rely on word of mouth to get the scoop; there are plenty of customer reviews and testimonials available online, so there's no need to be shy about asking around. Considering how readily available information such as dissects, end charts, and focus groups is, it may never be essential for an organization to solicit them in order to get notable evaluations. Seeing both positive and negative ideas, perspectives, and assessments is the task of thought appraisal. The majority of evaluation research has focused on narrative levels, such as distinguishing between positive and negative survey responses. However, tasks like mining item audits, theory-minded data extraction, and multi-viewpoint request observation and summing need evaluation at the phrase or even enunciation level. For example, how should it be configured to identify expressions of positive and negative concept if a solicitation seeing

framework appropriately addresses requests by considering persons' choices? The online cryo food research dataset is the basis for this evaluation. The data is all pertaining to surveys that provide an overview of several food transporters. People may use the transporters, which provide support while traveling, to post to Twitter. People often provide five different types of conclusions, such as "one point," "two points," "three points," and so on. People provide favorable feedback if they think the various forms of assistance supplied are worthwhile. Contradictory information is offered by individuals for terrible organizations. Twitter allows users to submit short messages called "tweets" on a regular basis and is part of the microblogging network that also rewards users for doing so. "Tweets" are brief communications that cannot exceed 140 characters. As a result of the nature of microblogging (brief, impassioned messages), users often resort to truncations, spelling mistakes, emojis, and other characters with fascinating meanings. Finally, we must not ignore the fact that evaluation is a problem with natural language processing. Since they are not addressed in natural language processing, it brings additional problems by contacting all of them: co reference targets, nullification management, and word meaning disambiguation. Whatever the case may be, it's also useful to understand that subjective evaluation is a basic limited NLP problem; this is because the framework doesn't have to fully comprehend the semantics of every sentence or report, just a few portions that are positive or negative and their objective components or subjects. That is why slant assessment provides a unique platform for NLP analysts to increase liberal indicators of ground across all NLP domains, which may have a huge objective impact. I use ML count to make assumptions and examine the aircraft dataset for conjecture in this report. I want to use this report to rally experts from various ML computations of NLP to collaborate on a smart project to solve the problem.

In little time at all, experts will have a far clearer picture of the whole scope, structure, and core difficulties of the problem. A plethora of novel models and systems are under consideration. Not only has the test not progressed recently, but it has also extended on a fundamental level. For the most part, prior fieldwork is necessary in order to rule out the potential of subjectivity in reports or sentences, which is inadequate for most real-world applications. Apps that are worth their salt often ask for incremental value and granular review. The report is produced in a format that is specifically designed for Twitter assessment exams utilizing an AI system,

which reflects the advancements in the industry. That is useful for developing a tailored framework for analyzing Twitter ideas. A straightforward way to dependably learn what everyone is up to in your gathering of partners is via social affiliations or long-range relational correspondence, both of which aid people in staying connected with their loved ones. Since you can be sure that your loved ones share many nebulous hobbies with you, easygoing affinities may also be used to find amusing and lovely items on the Internet. Brand confirmation and consumer agency are two outcomes of social engagement that advertisers want to achieve. Lengthy, easygoing communication drives a brand's voice and substance since it logically opens an association to new customers and gradually makes it obvious for current customers. For instance, a never-ending stream of Twitter users may think about a relationship while watching a news program and end up buying something or joining a club. An organization's chances of attracting and retaining customers are directly proportional to the degree to which its reputation is well-known. When people form relationships, it opens doors for them to communicate with one another via online platforms and apps. Examples of platforms that facilitate one-on-one communication include LinkedIn, Facebook, and Twitter. So far, the articulation casual association's relevance has been mostly a recent development dependent upon rapid alterations.

The goal of Estimation Analysis is to identify various types of experiences or data from a text segment using an approach that is associated with text analysis and distinctive language preparation. Analysts have done a plethora of study on this topic in the last several years. A great number of internet services, such as e come, are gradually gaining in popularity. Food delivery services offered online are one example. On the web, we may organize various foods according to our choice, and sometimes people will even provide surveys based on such foods. Typically, these surveys are discarded since they include unstructured data that cannot be used for anything further. This section of the study focuses on using the unstructured data to its full potential in order to get insight into the actions and reactions of customers on those digital platforms. There has been no perfect solution for hence requesting the enormous amount of tweets, leaving room for inquiry; yet, Twitter is an excellent tool for gathering consumer research and showing information cry audit dataset corporations. In order to manage and request tweets concerning transporter companies, this hypothesis focuses on establishing an alternative end portrayal strategy and distinguishing the introduction of diverse inclination of gathering ways.

## LITERATURE REVIEW

Additionally, data collection include text analysis. Regardless, there is a quality-control problem with the drug representation issue that stems from the typical data collection issue. Content data, which includes letters, words, or articulations, is overseen by message game plan applications, whereas most common data request applications handle numerical or apparent features. Transforming substance data into conventional numerical data and then realizing data representations is the most commonly accepted approach of applying normal data collection tactics to content demands. As an example, consider a food audit dataset where every word can be transformed into an attribute and every substance record into a vector of matching characteristics. This way, the file may display the occurrences of the words. The dimensionality of the updated, improved dataset is now going to be ridiculously high for representational tasks, however. Certainly, the articulations and lengthier grams may add up to more than 1,000 incontestable words in a food survey dataset. The substance procedure has several potential uses, including but not limited to: subject collection and evaluation request, language recognition verification, email coordinating, spam isolation, and topic collecting. The exponential growth of electronic and information technology has resulted in an enormous volume of electronic substances that is now impossible to rationally quantify. It has opened up new possibilities for the development of content request and other natural language processing methodologies but has also introduced new problems. In this way, content collection systems may coordinate monstrous electronic substance reports with managing development by using real or probabilistic calculations.

The substance request includes a feature decision system. The component assurance measure uses remember decision counts that are lower than the drug portrayal goal to choose features from the substance dataset. The subject request dataset may have its dimensionality reduced to an acceptable level by selecting the critical features for representation tasks. Twitter evaluation requests about food audit dataset have received very little attention, in contrast to substance presentation and end course of action. It is also lovely to come up with a different way to manage and boost the accuracy of the game plan, in addition to using significant evaluation request techniques to handle tweets about food audit examination. Lives lived online are programs and apps that let users to contribute and exchange information, thoughts, evaluations, etc. via a far-

reaching social communication site. The social media platform Twitter is also among the most popular. Twitter ranks third in overall and local use of long-range pleasant messaging sites, according to Alexa Internet Inc., an association that provides traffic statistics and assessment [1]. [2]. With 10 million users in the Philippines as of 2015, Twitter is likely the most popular long-form social media platform in the country. [2]. Data mining is the process of evaluating, collecting, cleaning, organizing, extracting, and expanding useful information from databases. Discomfortable areas, applications, subtleties, and information depictions taught about real applications are all part of a larger arrangement. In addition, these different pieces of data are sometimes referred to together as "information mining" [3].

The domains of speculation evaluation and emotional evaluation are quite similar, and the former may be considered a sub-region of the latter [4]. Examination mining focuses on the quality of studying the works, while hypothesis evaluation is more concerned with the substance's ultimate goal, independent of certainty or negativity; this is the only difference between the two approaches. There are three distinct tiers to End Analysis: report, sentence, and edge. To structure the whole report as an upward or downward trend is what the narrative level framework evaluation is all about. An approach to evaluating frameworks at the sentence level is to classify the ideas conveyed in each sentence. Finally, the conclusion of the specific edges or highlights of substances is what the point of view level assessment appraisal is hoping to witness. This analysis confirms that it is at the sentence level.

The aim of the 2015 evaluation by Sanket Sahu et al. titled "Twitter Sentiment Analysis" is to provide a consistently revamped method of representation and scoring. Using a spellcheck tally, they found that pre-planning was much better. Instead of painting tweets in a bad or positive light, they also offered a score system that allowed users to rate tweets according to their degree of essentiality or pessimism. Their classifiers of choice are maximum entropy and support vector machine. In the pre-dealing disconnected, they collected 60,000 tweets by removing external site links, usernames, RT for retweets, stop words, and emoji with words that were removed. They found that pre-dealing with is huge in brief opinionated material like tweets since it increased the organizing precision [5], and by the end of their assessment, 60,000 tweets had achieved an accuracy of 74.2%. Based on their evaluation, Meena Rambocas and Joao Gama concluded that the best way to find out what people think and feel about a



topic is to ask them directly. It combines feelings and perceptions instead of statistics. And this kind of information makes sense [6]. In 2012, Han et al. used an intriguing AI strategy called the K-closest neighbors (KNN) system. This system doesn't extract any highlights from the training dataset, but it does separate the report's similarity to its neighbors. Finding the k-closest tales and registering the measurements of the archives in different classes allows the KNN classifier to accumulate records to the class with the most neighbors for a given record d [7]. This has all the makings of an intuitively probable study by Thelwall et al. (2010) to determine if noteworthy events are usually associated with an improvement in idea quality [8]. An increase in negative thinking quality is often associated with well-known events, according to their results. In a similar vein, they verify that energy event peaks are more reliably positively rated than the period leading up to the peak. Using granular semantic ability in highlights for representation, Whitelaw et al. (2005) decouple their tendency approach method. assessments, such as "not appallingly sharp" or "overall eminent" [9]. In numerous task-independent semantic valid demands based on Appraisal Theory, an assessment gathering is attended to with a great deal of regard for various attributes. An innovative probabilistic presenting structure, the JST model depending on LDA, which perceives evaluation and selects topic continuously from the substance, is proposed by Lin et al. (2012) in an advanced report on the feeling branch of knowledge [10]. A topic is a probability distribution over words, and the LDA model is based on the notion that records include a mix of centers. A poorly thought-out concept, the JST merges an additional layer between the record and topic layers. As a result, JST is a four-tiered paradigm in which questions mark reports, themes mark end names, and words indicate assessment names and core interests. Similarly, Cai et al. (2010) construct a comprehensive evaluation mining framework that combines process for topic and thought affirmation [11]. In Turkish, there are usually not many reviews on the topic of doubt assessment using Twitter data. Considerations such as those that can identify positive and negative comments on news sites with an accuracy of 85% [12], social affair surveys of films based on positive and negative comments using a word reference-based strategy [13], and the categorization of tweets as positive, negative, or neutral using 2- and 3-gram models subject to word reference [14], investigating whether a correlation exists between the change in the cash-related exchange and the economic tweets of Twitter users [15], drawing on two clear informative datasets of Twitter and movie comments Using evaluations

based on word reference-based AI approaches, 75.2% and 85% of the time were successful on Twitter data sets, while 79.5% and 89% were successful on film audits [16]. A method that uses artificial intelligence to characterize these angles for certain types of cafés. Using a support vector machine (SVM) model to deduce the audit's conclusion bias from word frequency is the primary approach used in this work [17]. Customers' opinions on various businesses, products, and services may be found on review sites like Yelp and TripAdvisor, which have recently been shown to influence consumers' purchasing habits. The standard format for an online survey is free-form text followed by a 5-star rating. Review Rating Prediction is the challenge of predicting a customer's star rating for an item based on the customer's content survey for that item. It has lately been a popular, though challenging, topic in artificial intelligence [18]. Due to the exponential growth of e-commerce over the last two decades, food safety audits have become an important topic of discussion. Thus, food associations use the usual, very boring and dismal consumer analysis structures. When it comes time to gather consumer feedback tweets and conduct an emotional evaluation, Twitter data is a great resource. Using a dataset that included tweets for a critical howl food audit, we conducted a multi-class evaluation in this research. This philosophy begins with processes that are employed to clean the tweets in advance. Then, these tweets are addressed as vectors utilizing considerable learning thought to evaluate the articulation level. Four specific action systems—Random Forest, Logistic Regression, SVM, and Naïve Bayes—were used to complete the evaluation. The classifiers attempted to utilize the remaining 20% of the data after being configured to use 80% of the data.

The tweet inclination is the result of the test set. With the data in hand, we were able to compare and contrast the various data collection methods, and we were able to project the total number of predictions made by the Yelp food survey. In this investigation, we look at Twitter posts that are associated with food surveys and talk about evaluations in assumption examinations. The goal of this evaluation is to determine if tweets may be asked for as evidence of biased, biased-positive, or biased-negative thinking. Customers with an online life license may communicate with one another informally via correspondence objections, sharing and imparting knowledge, thoughts, and hypotheses. Notable transportation-related tweets are being compiled into a dataset to address consumer problems. The authors of this research developed a model using the classifiers Support Vector Machines (SVM), Naive

Bayes, Logistic Regression, and Random Forest to achieve limit via elucidation. Concerning their involvement with the organizations, the specialists collected tweets of nearby food audits. In order to better understand the preliminary result, the experts also picked the evaluation of a tweet based on whether it is positive, negative, or neutral, and provided quantitative and abstract evaluations in addition to the final evaluation. I have considered the preference evaluation depending on explorer contributions to transporter associations in this study. Our suggested method demonstrated that over-assessing procedures and component assurance are equally important for improving our outcomes. Our classifier setup calculations have been simplified thanks to feature decision methods, which have re-established the optimal subset of features. But without creating overfitting, it has reduced the lopsided class allocation in a large portion of our smaller datasets. In predicting occurrences that fall into the Positive, Negative, and Neutral categories, our findings provide convincing evidence that the suggested model has a high level of general accuracy. It is clear that some of the applied classifiers have outperformed the others. When tested on all datasets, algorithms including Random Forest, Logistic Regression, Support Vector Machines (SVM), and Naive Bayes demonstrated better security and a higher degree of requirements. Others, however, have shown an excellent degree of performance across all assessment estimates.

## RESEARCH METHODOLOGY

Tweets and messages may convey a wealth of information and opinions about people's lives and the world around them; the amount of detail they can convey is almost endless. Consequently, efforts are being undertaken to build a Twitter incline corpus in order to conduct research that will lead to a unique understanding of the transmission of evaluation in tweets and messages. The trademark language presents challenges when dealing with these laid-back substance classes because of the language used, which is incredibly laid-back and full of slang, off-base spellings, creative spelling and emphasis, new words, URLs, kind unequivocal expression, and shortened structures. In addition, data from electronic long-distance interpersonal contact, such as tweets, include rich coordinated information about the individuals involved in the discussion. For instance, Twitter preserves the details of who follows whom and retweets, in addition to the names included in

tweets, in order to provide the conversational data. Regarding the Yelp meal audit, this inquiry discusses and delivers evaluations in the suspicion assessment that Twitter gives. Finding out whether the tweets can be addressed in a favorable, bad, or reasonable light is the goal of this evaluation. Users of online communities are able to freely exchange information, ideas, and opinions using casual messaging platforms. Café tweets are becoming more and more popular and are being used as a dataset to assess customer concerns.

The authors of this research laid the groundwork for a model by explaining how they used the classifiers Logistic Regression, Random Forest, Support Vector Machines, and Naive Bayes. Tweets from customers on their experiences with local carriers at any restaurant were compiled by the experts. To further expedite the handling of the fundamental's aftermath, the experts also determined if a tweet's evaluation was positive, neutral, or negative, and provided quantitative and unique evaluations in addition to an end appraisal. Part A: Study Participants and Equipment According to this study, "Sentiment Analysis using Food Review Dataset" is the best title for the topic. In the field of natural language processing, this area is crucial for investigation. So far, I have evaluated the English language evaluation course using the specified and theoretical approach. For a critical learning model to work, you need a powerful computer with a variety of tools. An idea analyzer has provided the fundamental tool for this concept in a straightforward manner. Collecting Data (B) This delightful dataset includes the Yelp food audit dataset, which provides objective tweets based on customer surveys and ranks foods from one to five regions. In addition, the food quality explanation is the foundation upon which the tweets are sought. The request I need to fulfill requires just a handful of modifications, and this dataset doesn't need any cleaning exercises at the same time. Specifically, it is essential to separate the tweets into five categories: 1, 2, 3, 4, and 5. At that point, after consuming cry-sourced food, the customer will have completed the food audit survey. I must ascertain the finest and the most dreadful. From that point on, I must identify the core problems that led to a terrible trip. Think about just the tweets about bad meals for this.

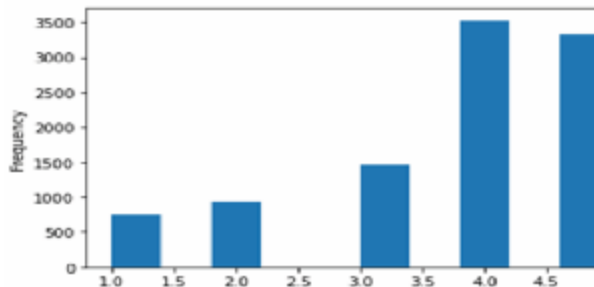


Fig. 1 Types of the sentiment in food review dataset

Presented in table 1 is a demonstration of the dataset. In the dataset, you may find several harsh tweets in the text section. After the material was pre-processed, it was moved into the process text section in an unaltered state. The related text's concept is contained in the notion section.

Table I Dataset demo table for yelp food review.

Review	Point
My wife took me here on my birthday for breakfast and it was excellent. The weather was perfect	5
Love the gyro plate. Rice is so good and I also dig their candy selection :)	4

Initial Stages Because raw data sometimes includes noise and duplication, or because the primary data isn't suitable for analysis methods, data pre-treatment is fundamental in data evaluation. Since content data is so different from numerical data, this is becoming more important in substance depiction; thus, we need to transform substance data into a separable design. Additionally, it is expected that data should be verified if the methods used for representation are controlled learning course techniques and if model assessments are included. Beyond that, real substance data typically has a plethora of syntax mistakes, restrictions, and images, which skews the conclusions of the algorithms. To illustrate the point, in informal network posts, individuals often use the symbols "withdraw" and "error" to signify important offers. This poses enormous challenges for AI figures to decipher, even if it is crucial for humans to understand that these terms represent something similar. A. D. Stemmer One such method of standardization is stemming. With the exception of tense, several different words might mean the same thing. We stem in order to standardize language and shorten the question. In terms of stemming

calculations, the Porter stemmer is among the most used.

T a b l e II Example of porter stemmer

pythoner	python
pythoning	python
pythonly	pythonli

Expression Regularized the python language The Python module provides comprehensive support for common enunciations in Python that are similar to Perl. If there is a problem with collecting or utilizing a standard explanation, the re module will issue the unique case re-error. In order to control frequent enunciations, I would address two crucial constraints. yet first, a minor yet important point: When utilized in the conventional verbalization, a number of characters would have a striking essentiality. (F) Terminate Take out Stop words are often used terms that online lists have been updated to disregard, whether they are noting fragments for skimming or remembering that retrieving them is the delayed result of an interesting request. Not to mention the enormous amount of time spent monitoring, I don't think these terms should eat up space in our database. I can clear things up enough by working with a list of terms that we think are stop words for this. An overview of stop words set aside in sixteen distinct vernaculars is provided by NLTK in Python. Their location in the nltk library may be found. Eighty percent of the data is prepared for training, while the remaining twenty percent is used for testing. A. Findings from the Experiment The datasets seem to be updated on a regular basis, and Random Forest, Naïve Bayes, SVM, and Logistic Regression were used as classifiers to assess the thought request. To put all the methods into action, the Python programming language is being used. To complete the model train, Sklearn is used. The estimation's portrayal settings were determined by observational testing of different objects with varied parameters. Ample precision measurements were used to examine the strategy. Table displays the delayed results of each dataset, with the dataset name shown under each plan results collecting. Changed numbers for each component assurance level reveal the outcomes and precision. You may see all the used calculations in the picture, together with their presentations, for your view. Based on the customer's rating, the meal review determines the sentiment. Each point is assigned a

star rating. Every review includes the customer's opinion on the food's quality. This was determined using four different algorithms. The following table contains all of the classification reports for the four methods. The collection includes 10,000 Yelp meal review records. Ten columns make up the dataset. The data utilized in the train portion is 80%. Twenty percent of the data is used in the test portion. The reviews are arranged in five star ratings.

## RESULTS AND DISCUSSION

The need for end assessment through Twitter has become an even more exciting area for research because it provides the public with dynamically groundbreaking assumptions, particularly through aviation organizations where people are both satisfied and confused enough to take their opinions to Twitter more often. When compared to other data sources, such as review sites or online diaries, where the data isn't just beneficial but also requires quantity, the data provided on Twitter is an unbelievable check. While sixty percent of the world's population uses Twitter—which, at first glance, may seem like a little amount—it really adds up to one hundred million users. On a daily basis, users post about one billion tweets, and that number continues to grow. This section provides an overview of the findings from the inquiry and examination. In addition, the discovery of the order report of used computations.

Table III. Accuracy table for multiclass classification.

Classifier	F1-score	Accuracy
Naïve Bayes	42.45	42.45%
Random Forest	48.80	48.80%
SVM	55.40	55.40%
Logistic Regression	51.80	51.80%

All four algorithms perform well in terms of accuracy, as seen in the accuracy table. The table demonstrates that SVM and logistic regression are two multiclass classification techniques that provide high accuracy for this dataset. For this dataset, the accuracy is low when using Naïve Bayes and Random Forest.

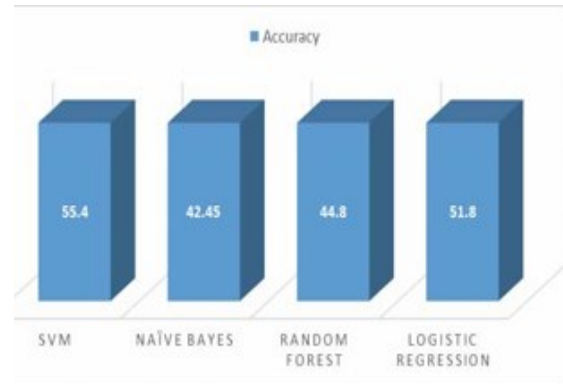


Fig. 4 Accuracy graph for sentiment analysis.

## Descriptive Analysis

Due to the abundance of visible information and the sheer volume of data that has to be explored, understood, and tested, end mining from sites like Twitter—which include lighthearted compositions—is necessary. To better evaluate Sentiment representation and gain semantic information from this region, a lot of research has been done. I provide a new method that provides a business model for classification using SVM, LR, NB, and RF estimates. Similarly, distinguish between the four calculations and provide the F1 score and accuracy. An Ensemble model with a solid foundational understudy as its foundation is the suggested solution. By retaining just the most informative characteristics and removing the noisy ones to create a robust vector of features, the component extraction of the module enhances the model's overall accuracy. After that, we sort the data from the Positive and Negative and objective stamps and imagine using the Ensemble techniques for course of action. Evidence suggests that the outfit process outperforms more conventional forms of data collection. Support vector machine (SVM) is a helpful guide that assigns more weight to poorly coordinated models after re-weighting them. Even with datasets that are skewed in one direction or another, SVM and LR provide exceptional hypothesis execution and capabilities. A continually summarized learning pointer count may be achieved with the help of this model, which combines point-based depiction with emoticons, retweet information, and more. There are a lot of potential extensions to this model.

## CONCLUSION AND FUTURE WORK

Researchers and experts from several fields have put a lot of consideration into the assessment strategy. A



number of philosophies have been developed for thought ordering in light of the many potential uses in commercial settings. Every sector is embracing the big data era and using data innovations to uncover untapped chances for improved partnerships. One such development is the speculation portrayal advancement, which can typically assemble consumer choices and provide a comprehensive understanding of customer analysis from raw Internet data. Twitter has been one of the most prominent places for displaying information exploration and end representation throughout the relational association phases. Looking at the introduction of several common inclination request ways and building a gathering strategy, which further enhances the idea portrayal execution, this hypothesis makes careful duties to this research zone. In addition, there is a wealth of useful information for improving food quality that comes from the evaluation of the collected tweets and the outcomes of the preliminary analyses. Finally, much like the consumers' Twitter habits, the uneven accuracy of the classifiers in various end classes reflect those habits. From recognizing minor speculative limits to seeing subjects' typical for these thoughts, evaluation mining has come a long way. Concurrently, my suggested method acquires the customer's findings and the subject's typical for assessments of this kind. As a result, the strategy makes them an essential issue and provides general facts and degree of the obvious customer aim, eliminating any guessing in the process. Responding to questions about the drivers of each stamped thought in a dataset and examining the idea's overall breadth are the goals of the suggested technique. To identify the hypothesis inquiry for the food audit dataset, I used standard ML grouping algorithms. SVMs provide very accurate results for evaluation tests.

## REFERENCES

- [1]. Tufts, P., Alexa Internet. 2007. Use of web usage trail data to identify relationships between browsable items. U.S. Patent 7,159,023. Top Sites in Philippines - Alexa. 2016. Retrieved from <http://www.alexatop.com/topsites/countries/PH> Aggarwal, C. C. 2015. Data mining: The textbook. Springer Publishing. IBM T.J. Watson Research Center, Yorktown Heights. DOI= 10.1007/978-3-319-14142-8 Cambria,
- [2]. E., Schuller, B., Xia, Y. and Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), pp.15-21.
- [3]. Sahu, S., Rout, S. K., & Mohanty, D. 2015. Twitter Sentiment Analysis -- A More Enhanced Way of Classification and Scoring. In *Proceedings of the 2015 IEEE International Symposium on Nanoelectronic and Information Systems*. DOI:10.1109/inis.2015.40
- [4]. Singh, B., & Singh, A. 2015. Strategies for Analyzing Quantitative Data in Research. Han, JiaWei, Micheline Kamber, and Jian Pei. 2012. in *Data Mining, Concepts and Techniques*, 364.
- [5]. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.
- [6]. Whitelaw, C., Garg, N., & Argamon, S. (2005, October). Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 625-631). ACM [10].
- [7]. Lin, C., He, Y., Everson, R., & Ruger, S. (2012). Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(6), 1134-1145.
- [8]. Cai, K., Spangler, S., Chen, Y., & Zhang, L. (2010). Leveraging sentiment analysis for topic detection. *Web Intelligence and Agent Systems*, 8(3), 291-302.
- [9]. Eroğul U. Sentiment Analysis in Turkish. MSc Thesis, Middle East Technical University, Ankara, Turkey, 2009.
- [10]. Vural AG, Cambazoğlu BB, Şenkuş P, Tokgöz ZO. "A frame work for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish". 27th International Symposium on Computer and Information Sciences, Paris, France, 3-4 October 2012.
- [11]. Meral M, Diri B. "Twitter üzerinde duygu analizi". IEEE 22. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Trabzon, Türkiye, 23-25 Nisan 2014.
- [12]. Şimşek M, Özdemir S. "Analysis of the relation between Turkish twitter messages and stock market index". 6th International Conference on Application of Information and Communication Technologies (AICT), Tbilisi, Georgia, 17- 19 October 2012.
- [13]. Türkmenoğlu C, Tantuş AC. "Sentiment analysis in Turkish media". Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM '14), Beijing, China, 21-26 June 2014.
- [14]. Yu, Boya, et al. "Identifying restaurant features via sentiment analysis on yelp reviews." *arXiv preprint arXiv: 1709.08698* (2017).
- [15]. Luo, Y., & Xu, X. (2019). Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp. *Sustainability*, 11(19), 5254