International Journal for Modern Trends in Science and Technology, 9(02): 60-63, 2023 Copyright © 2023 International Journal for Modern Trends in Science and Technology ISSN: 2455-3778 online DOI: https://doi.org/10.46501/IJMTST0902010

Available online at: http://www.ijmtst.com/vol9issue02.html



Predicting the Reviews of The Restaurant using Natural Language Processing Technique

Mr.B.Venugopal¹ | G.Aasritha² | CH.Venkata.Jayanth Kumar² | E.R.Chandra Mouli² | B.Jaya Kumar²

¹Associate Professor, Department of CSE, NRI Institute of Technology, India ²B.Tech Student, Department of CSE, NRI Institute of Technology, India

To Cite this Article

Mr.B.Venugopal, G.Aasritha, CH.Venkata.Jayanth Kumar, E.R.Chandra Mouli, B.Jaya Kumar. Research on Predicting the Reviews of The Restaurant using Natural Language Processing Technique. International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 60-63. <u>https://doi.org/10.46501/IJMTST0902010</u>

Article Info

Received: 30 December 2022; Accepted: 01 February 2023; Published: 03 February 2023.

ABSTRACT

In this age of the internet, an incredible quantity of data is now being sent through the network. It is now normal practice for consumers to research goods and services that they are considering purchasing on the internet since the breadth of material available on the internet encompasses both subjective opinion and objective facts. However, due to the fact that a significant portion of information is only available in the form of text fragments and does not include any sort of numerical scales, it is difficult to categorize their assessment in an effective manner without reading the complete text. In this section, we will concentrate on harvesting scored ratings from snippets of text found on the internet, and we will recommend a variety of experiments in an effort to increase the overall quality of a classifier.

KEY WORDS: Sentiment Analysis, Ensemble Learning

1. INTRODUCTION

Getting feedback from clients regarding the quality of the services that a company provides is something that many companies strive to do in order to improve and increase their revenues. Restaurant-goers could be interested in learning from the experiences of others based on a range of characteristics, such as the quality of the cuisine, the service, the atmosphere, discounts, and how worthwhile the restaurant is. Yelp users are able to publish their own reviews, ratings, and comments on local businesses and services, or they may just respond to the reviews of others. From one person's point of view, bad reviews may have an impact on future consumers' ability to make choices; for example, a potential customer may decide to terminate a service and may convince other potential customers to do the same. There is no shortage of research conducted on the Yelp dataset from a variety of perspectives, which reveals useful information on a broad range of topics such as the effect of promotion strategies [1], the benefit of retrieving knowledge from implicit user feedback [2], or the important role of local reviewers [3]. Some of these topics include: [a] [b] [c] [d] [e] [f] [g] [h] I [j] [k] The failure to capture quick changes in knowledge is a shortcoming shared by the vast majority of research. As a result, the results can only be considered reliable when viewed in the context of a static perspective on business activity. In the actual world, a changing data source like Yelp could be better represented using a dynamic method to correctly reflect continual changes in business operations. This would be the case since Yelp is always evolving. One example of such a dynamic strategy is known as incremental learning, and it has the capacity to learn a new idea without having to undergo retraining on the whole dataset. Using a method that is based on incremental learning, the issue that has to be answered is how to quantify the ways in which consumers and companies are impacted, as well as how evaluations of businesses shift in response to nal recent input.

2. LITERATURE SURVEY

The majority of research on sentiment analysis relies on machine learning algorithms, the primary purpose of which is to determine whether a given text is in favor of or opposed to something and to determine the polarity of language. In this chapter, we will discuss some of the research that has been done on the subject, which has contributed to our in-depth understanding of it. P. Pang, L. Lee, S. Vaithyanathan et al [8] They pioneered the field of sentiment analysis with their work. Their primary objective was to categorize texts not just according to the subject matter discussed, but also according to the general tone of the writing. For example, they wanted to determine if a movie review was good or negative. They use a machine learning algorithm on a database of movie reviews, which ultimately leads to the conclusion that their algorithms are superior to those created by humans. The Naive Bayes, Maximum Entropy, and Support Vector Machines machine learning algorithms are the ones that are used by this company. They also reach the conclusion, after considering a number of different elements, that categorization of emotion is an extremely difficult task. They demonstrate that supervised machine learning methods are the fundamental building blocks for sentiment analysis. P. Pang, L. Lee et al [9] Finding out what people are thinking or what they anticipate based on enormous amounts of data collected has always been an important step. The development of the sector of social media has led to an increase in the availability of data that is rich in the resources available to form opinions. We are gaining a better understanding of what individuals can accomplish and their perspectives on the issue thanks to additional resources such as blogs, review sites, communications, and so on.

The rapid growth in the amount of work being done in the area of data mining and sentiment extraction is concerned with the use of computer capacity to tackle the issue of opinion mining or the presence of subjectivity in text. As a result, a wide variety of brand new systems are being developed on the basis of a variety of languages and instructions. These systems are able to deal directly with opinion mining as the first class object, and direct response or live research is also becoming an area of interest. 13 They conduct a survey in which the topics of methodology and techniques that are used in direct response opinion mining are covered. Some of these methodologies and approaches are more beneficial than others. They are concentrating on developing functions that may handle new problems that have arisen in applications using sentiment analysis. They also made a comparison between these innovative methods and the more conventional, fact-based analysis that is currently in use. E. Loper, S. Bird et al [10] Natural Language Toolkit (NLTK) is a library that includes many different software modules, a big amount of structured files, several statistical fun<mark>ction</mark>s, ready-to-use machine learning classifiers, multiple tutorials, problem sets, and computational linguistics courseware, among other things. The primary objective of the Natural Language Toolkit (NLTK) is to carry out natural language processing, often known as doing analysis on data pertaining to human language. Corpora, which are essential for the training of classifiers, are made available by NLTK. New components are built by developers, some of which are designed to replace older ones. As a current programs become consequence, more organized, and the results that they provide are more complex. H. Wang, D. Can, F. Bar, S. Narayana et al [11] They were the academics that came up with the idea for a system that could analyze public reactions in the 2012 presidential elections in the United States in real time. They gather the answers from Twitter, which is a platform for micro blogging. People express their ideas, thoughts, and opinions about any issue that is currently trending on Twitter, which is one of the social networking sites. People's reactions to election candidates in the United States on Twitter collected a vast quantity of data, which helps to construct a sentiment for each candidate and also made a forecast of who would win. Twitter is an online social networking service. The reactions of individuals on Twitter are connected to the outcomes of the election, which results in the formation of a causal relationship between the two. They investigate the role that sentiment analysis plays in these public events as well. In addition to this, they demonstrate that live sentiment analysis may be finished relatively quickly in contrast to conventional content analysis, which might take several days or even a few weeks to finish. The system that they exhibited does an analysis of the sentiment of all of the Twitter data pertaining to the election, candidates, promotions, and other related topics, and it delivers findings at a constant rate of 14. It provides the media, legislators, and academics with a whole new method that is both timely and successful, and it is entirely based on public opinion. O. Almatrafi, S. Parack, B. Chavan et al [12] Researchers like them came up with the idea of a system that is dependent on location. They claim that Natural Language Processing (NLP) and machine learning techniques are used in sentiment analysis in order to extract a sentiment from a text unit that originates from a certain area. They investigate several applications of location-based sentiment analysis by using a data source from which information can be readily collected from a variety of places, and they do this while doing their research. There is a feature on Twitter called "tweet location," and because this information can be accessed rather simply by a script, it is possible to gather data (tweets) from a certain place in order to find trends and patterns. Within the scope of their investigation, they focus on the general elections in India in 2014. They do mining on 600,000 tweets that were gathered for two political parties over the course of seven days. In order to construct a classifier that is capable of categorizing tweets as either positive or negative, they use supervised machine learning techniques, such as the Naive Bayes algorithm. They determine the ideas and views of users in regard to these two political parties in a variety of regions, and then they plot their findings on a map of India by using a Python package.

3. PROPOSED SYSTEM

The purpose of our proposed system is to utilize methods from natural language processing in order to categorize a collection of restaurant reviews according

to the number of stars that were awarded to each review. We classify each review on a scale from one to five stars using a classification system based on maximum entropy. We implement a collection of characteristics that we consider to be significant to the sentiment that is stated in reviews, and we examine the influence that these features have on performance. This provides us with insights into what works, as well as why sentiment classification may be so challenging. We investigate the ways in which the tone of a review might influence the degree to which it adheres to a certain linguistic model. We try out a variety of models of emotional expression that are linguistically informed, and we use the findings of these experiments to further enhance the effectiveness of our classifier. We investigate how tagging words with their parts of speech affects our ability to deduce the speaker's feelings. We tried out a variety of approaches to preparing the data in our experiments. The reviews are unstructured in terms of the user input, which means that they can look like anything from a paragraph of well-formatted text to a jumble of seemingly unrelated words to a run-on sentence with no apparent regard for grammar or punctuation. This is because the reviews are not moderated. Our first run through the data consisted of just tokenizing the testimonials on the basis of whitespace and treating each token as a unigram. However, we were able to increase speed by eliminating punctuation in addition to the whitespace and transforming all of the letters to lowercase. Because of this, we evaluate all instances of "good," "Good," and "good." as being equivalent to one another. This results in increased predictive power for any test set review that contains any of these three forms. Before converting into the unigram, stemming was also performed, which means that the multiple forms (tenses, verbs) of the words were eliminated and handled as if they were a single word. In order to achieve a higher level of precision and exclude terms that are used seldom, a threshold is applied once the matrix has been constructed. Therefore, relevant unigrams and bigrams that are happening more than the threshold number of times are both included in our matrix.

4. RESULTS



Figure1: Review Page



Figure2: Accuracy

5. CONCLUSION

While the proliferation of social media has provided unprecedented chances for individuals to openly voice their ideas, the Internet's ability to analyze this data is severely limited. The need to grasp the worries of the populace in real time has also increased fast. Due to the quick and widespread dissemination of information made possible by social media, many topics gain unexpected traction as a result of the users' own words of mouth in the form of online comments and reactions. The decision makers and individuals have not yet grasped that sentiment analysis may help them make sense of this mass communication and communicate rationally with thousands of others. This paper provides a concise overview of sentiment analysis, outlining its definition, its early history, its problems, its basic concepts and their mathematical treatment, its sentiment and subjectivity classification, which is made up of opinion mining and summarization, and its application in real-world business settings. The use of mathematical techniques to analyze sentiment is investigated, along with other approaches such as "sentiment analysis as text classification problem" and "sentiment analysis as feature classification." With the use of social media sentiment research, we also present the monetary effects that this field has on people,

communities, and businesses. We have proposed an NLP algorithm and demonstrated its performance for learning sentiment analysis. However the assumption that we know the set of all possible values for features when extracting features is not always practical.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Ariyasriwatana, W., Buente, W., Oshiro, M., & Streveler, D. (2014). Categorizing health-related cues to action: using Yelp reviews of restaurants in Hawaii. New Review of Hypermedia and Multimedia, 20(4), 317-340.
- [2] Byers, J. W., Mitzenmacher, M., & Zervas, G. (2012, June). The groupon effect on yelp ratings: a root cause analysis. In Proceedings of the 13th ACM conference on electronic commerce (pp. 248-265). ACM.
- [3] Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., & Bevan, J. L. (2012). Why people use Yelp. com: An exploration of uses and gratifications. Computers in Human Behavior, 28(6), 2274-2279.
- [4] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013, July). What yelp fake review filter might be doing?. In ICWSM. 6
- [5] dos Santos, C. N., & Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In COLING (pp. 69-78).
- [6] Mullen, T., & Collier, N. (2004, July). Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In EMNLP (Vol. 4, pp. 412-418).
- [7] Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014, August). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 437-442). Dublin, Ireland: Association for Computational Linguistics and Dublin City University.
- [8] Huang, J., Rogers, S., & Joo, E. (2014). Improving restaurants by extracting subtopics from yelp reviews. iConference 2014 (Social Media Expo).
- [9] Shalev-Shwartz, S., Singer, Y., Srebro, N., & Cotter, A. (2011). Pegasos: Primal estimated sub-gradient solver for svm. Mathematical programming, 127(1), 3-30.
- [10] Saif, Hassan, et al. "On stopwords, filtering and data sparsity for sentiment analysis of Twitter." (2014): 810-817