International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 6, pp: 740-743 www.ijaem.net ISSN: 2395-5252

Comparison of Machine Learning Techniques for Sentimental Analysis on Restaurant Reviews

[1]Haja Sharieff, [2]T. Sindhu, [3]L. SaiRamesh
[1]*Ramanujan Computing Centre, Anna University, Chennai, Tamilnadu, India*
[2,3]*Department of IST,Anna University, Chennai, Tamilnadu, India.*
Corresponding Author: Haja Sharieff

Date of Submission: 27-09-2020 Date of Acceptance: 29-09-2020

ABSTRACT: Opinion mining or sentiment analysis analyses the text written in a natural language about a topic and classify them as positive negative or neutral based on the human's sentiments, emotion, opinions expressed in it. Sentiment analysis of customer reviews has a crucial impact on a business's development strategy. Despite the fact that a repository of reviews evolves over time, sentiment analysis often relies on offline solutions where training data is collected before the model is built.In this project will collect and analyse the reviews about the restaurants and helps us understand the people reaction and opinion towards that restaurant. Through this analysis we can generate the analysis report regarding the opinion of the people about a particular restaurant.The purpose of this analysis is to build a prediction model to predict whether a review on the restaurant is positive or negative. To do so, we will work on Restaurant Review dataset, we will load it into predictive algorithms Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression. In the end, we hope to find a "best" model for predicting the review's sentiment.

KEYWORDS: Opinion mining, Sentiment analysis, Aspect extraction, Multinomial Naïve Bayes.

I. INTRODUCTION

In this paper, restaurant recommendation based on the review analysis is proposed. The recommendation is carried out based on following queries. What makes a good restaurant? What are the major concerns of customers for a great meal? Common knowledge may give general answers like delicious food, great services or pleasant environments, but they might not be true for different types of restaurants. In this project, we are going to unveil those essential features behind all kinds of restaurants via sentiment analysis on data.

Businesses often want to know how customers think about the quality of their services in order to improve and make more profits. Restaurant goers may want to learn from others' experience using a variety of criteria such as food quality, service, ambience, discounts and worthiness. Users may post their reviews and ratings on businesses and services or simply express their thoughts on other reviews. Bad (negative) reviews from one's perspective may have an effect on potential customers in making decisions, e.g., a potential customer may cancel a service and persuade other do the same.

.

II. LITERATURE SURVEY

In the research work [1], new hybrid classification method is proposed based on coupling classification methods using arcing classifier and their performances are analyzed in terms of accuracy. A Classifier ensemble was designed using Naïve Bayes (NB), Support Vector Machine (SVM) and Genetic Algorithm (GA). In [2], the user interests are captured based on their feedback received implicitly from the user's activities. The user's searching behaviour is ranked according to their usages.

In the paper [3], they have introduced a machine learning based method to characterize such aspects for particular types of restaurants. The main approach used in this paper is to use a support vector machine (SVM) model to decipher the sentiment tendency of each review from word frequency.Sentiment analysis of customer reviews has a crucial impact on a business's development strategy. Despite the fact that a repository of reviews evolves over time, sentiment analysis often relies on offline solutions where training data is collected before the model is built. If we want to avoid retraining the entire model from time to time, incremental learning becomes the best alternative solution for this task [4].

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 6, pp: 740-743 www.ijaem.net ISSN: 2395-5252

In [5], the sentiment analysis is analysed based on the opinion of the review word. Semi-supervised word alignment model is used to extract the real opinion of the reviews. In [6], the opinion provided in the tweets are extracted and predict the future event based on the existed tweets. Another work in [7], correlate the multiple and group them to identify the relevant reviews based on the opinion words. In [8], similar users are group by using K-means clustering algorithm. It is used to cluster the users based on the keywords they are mentioned in the review. The survey is discussed in [9] to show the challenges in personalization of web search.

III. PROPOSED METHODOLOGY

To build a model to predict if review is positive or negative, to do so, we have worked on the Restaurant Review dataset, we have loaded it into predictive algorithms Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression. In the end, we hope to find a "best" model for predicting the review's sentiment. following steps are performed.

Importing Dataset

Restaurant_Reviews.tsv is a dataset from Kaggle datasets which consists of 1000 reviews on a restaurant. We have imported the libraries NumPy and pandas for the project. NumPy is not another programming language but a Python extension module. It provides fast and efficient operations on arrays of homogeneous data. NumPy extends python into a high-level language for manipulating numerical data.

Pandas is an open-source, BSD- licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Pre-processing Dataset

Each review undergoes through a pre-processing step, where all the vague information is removed like removing the stop words, numeric and special characters. We have imported the library NLTK and from NLTK we have imported stop words and Porter Stemmer for removing the stop words, numeric and special characters.

NLTK(Natural Language Toolkit) is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries.

Vectorization

A way to represent text data for machine learning algorithm and the bag-of-words model helps us to achieve that task. The bag-of-words model is simple to understand and implement. It is a way of extracting features from the text for use in machine learning algorithms. The process of converting NLP text into numbers is called vectorization in ML. From the cleaned dataset, potential features are extracted and are converted to numerical format. The vectorization techniques are used to convert textual data to numerical format. Using vectorization, a matrix is created where each column represents a feature and each row represents an individual review.

sklearn.model_selection import train_test_ split .It Split arrays or matrices into random train and test subsets Quick utility that wraps input validation and next (ShuffleSplit().split(X, y)) and application to input data into a single call for splitting (and optionally subsampling) data in a one-liner.

Training and Classification

Further the data is split into training and testing set using Cross Validation technique. This data is used as input to classification algorithm.

Classification Algorithms:

Algorithms like Decision tree, Support Vector Machine, Logistic Regression, Naive Bayes were implemented and on comparing the evaluation metrics two of the algorithms gave better predictions than others.

1. Multinomial Naive Bayes
2. Bernoulli Naive Bayes
3. Logistic Regression

IV. EXPERIMENTAL RESULT AND ANALYSIS

In this study, we have made an attempt to classify sentiment analysis for restaurant reviews using machine learning techniques. Three algorithms namely Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regrssion are implemented.

Evaluation metrics used here are accuracy, precision and recall as shown in Figure 2, 3, 4 and 5.

Using Multinomial Naive Bayes,

- Accuracy of prediction is 77.67%.
- Precision of prediction is 0.78.
- Recall of prediction is 0.77.

Using Bernoulli Naive Bayes,

- Accuracy of prediction is 77.0%.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 6, pp: 740-743 www.ijaem.net ISSN: 2395-5252

- Precision of prediction is 0.76.
- Recall of prediction is 0.78.

Using Logistic Regression,

- Accuracy of prediction is 76.67%.
- Precision of prediction is 0.8.
- Recall of prediction is 0.71.

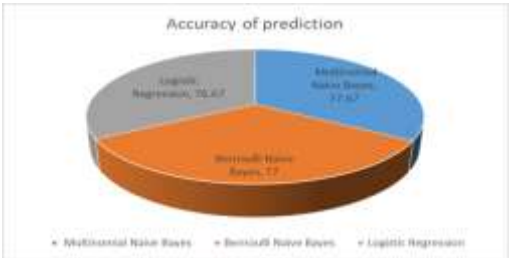

Figure 2. Accuracy of Prediction

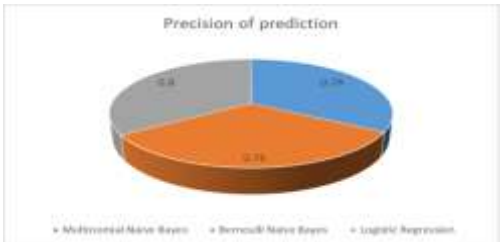

Figure 3. Precision of Prediction

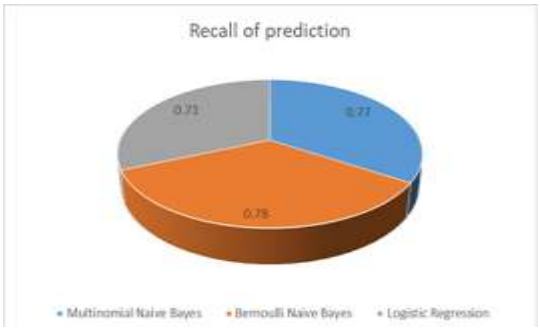

Figure 4. Recall of Prediction

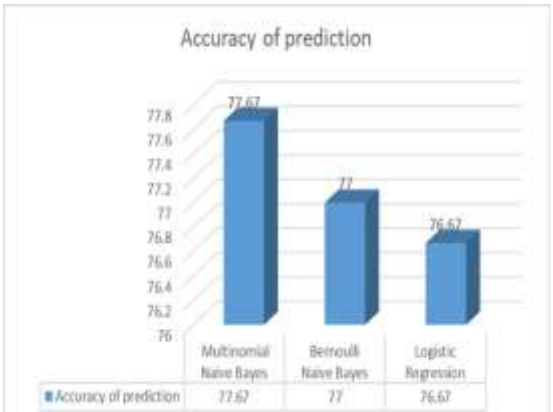

Figure 5. Accuracy of Prediction

From the above results, Multinomial Naive Bayes is slightly better method compared to Bernoulli Naive Bayes and Logistic Regression, with 77.67% accuracy which means the model built for the prediction of sentiment of the restaurant review gives 77.67% right prediction.

V. CONCLUSION

In this work, we have collected and analyse the reviews about the restaurants and helps us understand the people reaction and opinion towards that restaurant. Through this analysis we can generate the analysis report regarding the opinion of the people about a particular restaurant. Through this analysis, many restaurants owners and chains a gets to know about the public opinion about their brand or their food quality and this help them to improve or analyse their food quality according to public opinion analyses. Multinomial Naive Bayes is slightly better method compared to Bernoulli Naive Bayes and Logistic Regression, with 77.67% accuracy which means the model built for the prediction of sentiment of the restaurant review gives 77.67% right prediction. In marketing field, many brands use it to develop their strategies, to understand customers' feelings towards the food quality and the services of the restaurant.

REFERENCES

[1]. Govindarajan, M. (2014). Sentimentanalysis of restaurant reviews using hybrid classification method. International Journal of Soft Computing and Artificial Intelligence, 2(1), 17-23.

[2]. Ramesh, L. S., Ganapathy, S., Bhuvaneshwari, R., Kulothungan, K., Pandiyaraju, V., & Kannan, A. (2015). Prediction of user interests for providing relevant information using relevance feedback and re-ranking. International Journal of Intelligent Information Technologies (IJIIT), 11(4), 55-71

[3]. Yu, B., Zhou, J., Zhang, Y., & Cao, Y. (2017). Identifying restaurant features via sentiment analysis on yelp reviews. arXiv preprint arXiv:1709.08698.

[4]. Doan, T., & Kalita, J. (2016, December). Sentiment analysis of restaurant reviews on yelp with incremental learning. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 697-700). IEEE.

[5]. Sadhana, S. A., SaiRamesh, L., Sabena, S., Ganapathy, S., & Kannan, A. (2017, February). Mining target opinions from online reviews using semi-supervised word

alignment model. In 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM) (pp. 196-200). IEEE.

[6]. Sulthana, A. R., Jaithunbi, A. K., & Ramesh, L. S. (2018). Sentiment analysis in twitter data using data analytic techniques for predictive modelling. In J. Phys. Conf. Ser (Vol. 1000, No. 1).

[7]. Sabena, S., Kalaiselvi, S., Anusha, B., & Ramesh, L. S. (2016). An Multi-Label Classification with Label Correlation. Asian Journal of Research in Social Sciences and Humanities, 6(9), 373-386.

[8]. Selvakumar, K., Ramesh, L. S., & Kannan, A. (2015). Enhanced K-means clustering algorithm for evolving user groups. Indian Journal of Science and Technology, 8(24)

[9]. Selvakumar, K., &Sendhilkumar, S. (2011, December). Challenges and recent trends in personalized Web search: A survey. In 2011 Third International Conference on Advanced Computing (pp. 333-339). IEEE.

.

.