IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Customer's Sentiments based Review Prediction

Rupesh Sonkamble, Arati Manjaramkar

M. Tech. Information Technology,
Head of Dept.
Information Technology
Shree Guru Gobind Singhji Institute of Engineering & Technology
Nanded, India

Abstract: As the speed of innovation increases at an accelerating rate, Individuals' ways of sharing their views on different websites is also expanding. On social media sites like Facebook, Twitter, and Yelp, there are various reviews and opinions. Scores are normally provided on a range of 1 to 5 stars. Assessment of opinions through textual information analysis has shown an important impact in analytic research because this offers useful options to emotions mining. Review is an evaluation of a product or service by someone who has used the product or service or has experience with it. The ranking of any e commerce site is heavily influenced by the opinions of its users. The purpose of this paper is to explain how a Machine Learning (ML) set of rules works on Yelp's database to evaluate, anticipate, and advocate brands. With the sentimental analysis algorithm, we analyzed Support Vector Machines, K- Nearest Neighbor, Multilayer Perceptron classifiers, Naive Bayes, , Decision Tree and Random Forest. The best result of the multi-layer perceptron classifier is 93.40 percent.

***Index Terms* - Machine Learning, Sentimental Analysis, Opinion Mining, Natural Language Processing, Review rating prediction.**

I. INTRODUCTION

In this age of modernization, we are constantly in search of some idea that will allow us to save time; reduce the task's complexity by substituting physical production. Sentiment analysis is a special technique which helps us to save time while still accomplishing our task. Rather than detecting emotions, data analysts use the phrase "sentiment analysis." Opinion mining, also known as sentiment mining, is a form of natural language processing that identifies the customer's feelings, expressions, and views about a particular object, film, condition etc. User reviews are an important part of internet applications like Food panda, Amazon, Swiggy, Yelp, Zomato and others, in a place where customers share their opinions on companies, products, and services between free-from text reviews and star ratings, typically out of five. These reviews serve as "digital word-of-mouth" & as a consumer search component for the same products available in the market. According to study, they have a huge influence on consumer purchase decisions, as well as brand sales and revenue. Only extracting features will not solve the problem; for the best recognition rate, we must also identify the suitable classifier. The Review Rating Prediction Issue is classified as a multi-class classification problem in Machine Learning, with the class labels being the star ratings. On the raw data, we used Data Cleaning, Data transformation, and Feature Extraction, as well as a variety of machine learning algorithms to predict user reviews based on star ratings. The following is a breakdown of the paper's structure. The related study of predicting rating based on user attitudes described in Section II . Section III discusses the Yelp data set, while Section IV discusses machine learning and its proficiency. Section V explains the process flow. In Section VI, the results of the experiments are discussed. Section VII discusses future efforts and the conclusion.

II.Related Work

Micro blogging is an increasingly popular broadcasting platform among the Internet population these days. Every day, people share their thoughts and views about a wide range of issues on micro blogging websites including products, media, organizations, and so on. In estimation systems, perception mining, and other applications, sentiment evaluation is important. Users on Facebook, one of the micro blogging websites, are limited to 140 characters. For data recovery purposes, Twitter often offers a programmer-friendly streaming API, allowing the researcher to check several users for real-time tweets. This far has been the case. Machine learning techniques have proved to be extremely successful, delivering accurate outcomes. Any method's efficacy is essentially determined by the order. The lexical technique is a ready-to-use approach that does not require any prior expertise or instruction. Despite the fact that machine learning involves a well-made classifier, It wishes a big variety of education information units and regression testing before it can be deployed. On twitter data sets, The accuracy of sentiment analysis using class-two naive bayes is 84%.

In social networking sites, the way people share their views, assessments, or feelings about items or companies has changed drastically. Among other social media sites, one of the top social websites is twitter, which offers customers to share their daily activities. As a positive or negative tweet emotion, the authors illustrate & provide machine learning system functionality. Data collection and storage of tweets using the Twitter API and the top of the IPL hashtags (# IPL 2016 and #IPL 9) which should be accomplished. The final result was calculated using the relevant sample dataset and Random Forest techniques. As an output parameter for each tweet, A binary-class value of '0' or '1' is used to represent both positive and negative views. The evaluation accuracy of the proposed technique is 81.69 percent. The method of analyzing languages to analyze if a part of a review has subjective data & the type of subjective data it expresses is known as sentiment analysis. The authors used big data to examine vast volumes of tweets and calculated the polarity of letters, sentences, and whole articles. Author uses linear regression to investigate the connection between independent variables labelled X and a scalar dependent variable Y. This technique of information analytics is greater green than the use of vector machines and naive bayes. The method's precision is 85.23 percent, which is a major improvement over SVM. The authors use 10-fold cross validation to increase the system's accuracy.

L. Kao and Y. Huang [5] suggests a data mining method for evaluating the association between the sentiments of fan page users and client purchasing acts to predict past client purchasing activity. Businesses create their own fan pages to advertise their products and broadcast advertisements. Followers to the fan page often post their opinions on the website, which helps to spread customer opinion. Because all of these viewpoints will be disseminated across the social network to every nook and cranny, some business analysts wonder whether they will help the sale of products. The first phase of the proposed method is to determine a single fan's sentiment rating based on their messages. The term "public sentiment" refers to a summary of all of the fans' emotions. To discover the series like "if cutting-edge message of high-quality sentiment has a tendency to increase, the amount of merchandise bought will enhance over the subsequent days," The link between public sentiment and purchase volume is then determined using the inter-transaction affiliation rule mining approach.

F. Ren and Y. Wu [6] focuses on the difficult task of Estimating personal sentiments on subjects that they haven't directly addressed before, which the authors refer to as user topic opinion estimation. The Twitter dataset tells methods outperform new filtering methods, as evidenced by research findings. The author conducts tests to assess the proposed ScTcMF technique. Social context and topical context are helpful in enhancing the quality of customer sentiment analysis.

R. Chakraborty et al. [7] show that outlining tweets related to news necessitates getting the better of unique challenges so as to achieve different goals, including securing equality, relevance, and reporting. Nowadays tweet average strategies primarily deal with events attempting to create a collection of relevant tweets before managing the criteria for diversity and coverage sequentially via a specific report. The authors discovered that the way the summary's goals are presented has a significant impact on the summary's quality. Initial research revealed that managing tweet diversity correctly should improve the life of averaged tweets and thus the average performance. The counseled approach has been established for political news stories and can also be used to obtain summary tweets for specific news classes.

Consumer satisfaction polls, in which questions are explicitly designed to gather customer input on particular goods or services, have traditionally been used to test public opinion. J. Zhu et al. [8]. Present aspect-primarily based totally opinion polling as opposed to unlabeled free-shape consumer conceptual comments and not using requests for solutions to questions. Then a multi-aspect bootstrapping device should be purchased. A new methodology called aspect-based segmentation method suggests segmenting multi-element sentences into a couple of single-element bases as widespread devices for opinion polling. The proposed method of opinion polling focused on aspects achieved. In the experimental findings, there was 75.5 percent accuracy on real Chinese restaurant reviews.

On sites like Amazon, Netflix, and Yelp, there are a number of reviews and ratings. Scores are typically given on a scale of one to five stars. Reviews are a type of free-form text that consists of many words. Textual sentiment analysis has been described as a critical component of analytics research because it offers useful options for opinion mining. We may recommend new products, films, and restaurants to a person based on their feedback and ratings. Through observing similar customers and developing feedback, recommenders are able to fit consumer actions. Developing suggestions. From the point of view of a user's responsibility of the film and interests, a scheme with sentiment tags coupled with regular recommendations appears to be a creative and rational solution. Random Forests, Unigram, Bigram, Trigram, Naive Bayes, Support Vector Machines are among the methods used by B. S. S. Govind et al. [9]. In comparison to traditional techniques, the authors use the Random forest approach to significantly boost sentiment mining. Better data stemming and pre processing processes led to a decrease in root-mean-squared-error (RMSE). The recommendation system's calculation time is greatly decreased when Spark is used.

J. Wang [10] performed a sentiment analysis based on Yelp user feedback. Author uses user input to build predictive fashions for ratings in a given dataset. The author finds a 5 or 4 star Yelp rating to be positive and a 3, 2 or 1 rating to be negative. Author handed-down supervised learning algorithms such as SVM, perceptron learning algorithm, and Naive Bayes to see emotions as shown in the Yelp system for sentiment prediction. The features from the Yelp user reviews material was extracted using a number of language systems.

III. Dataset

Yelp [12] is the biggest online comment forum in the United States. It includes businesses in the restaurant, grocery, hotel, and hospitality sectors. Yelp allowed people to review and rate local businesses in the same way they would online for products. Yelp's website and mobile applications have a diverse range of suggestions. "Yelp's Best: City" is one of Yelp's most well-known groups. Yelp ranks businesses based on Yelp user feedback and ratings, and chooses the most popular places to visit in a town in this category. When it comes to discovering a company, most users rely solely on the average star rating. Although it offers a clear interpretation of a person's opinion of a company, it involves a lot of different types of data, such as features, customers, and the review text itself.

Customers can leave comments, check ins, ratings, and other information about any company they like or dislike on the Yelp website. People often "Yelp" the location to which they are about to fly, or use Yelp's assistance to decide which businesses are doing well. Unlike many other appraisal and suggestion systems, Yelp builds its own social network with a graph of people connected to his or her peers, rather than relying on common social networks like Facebook or LinkedIn. Users may assign a business a star rating from 1 to 5, as well as write a text review to back up the number ranking. Those ratings are a great guide for users searching for local businesses, helping them to find out which one is right for them and making Yelp a reliable source of feedback.Every enterprise has a popular ranking, that's clearly the sum of the big name rankings for all the evaluations the organization has evaluated. Users also can vote in choose of opinions written with the aid of using different users, with votes ranging from helpful to cool to funny. User data, company data, user-to-business access history, user scoring, and responses are all included in the Yelp dataset.Yelp has made a portion of the data accessible to the scientific and educational communities by organizing competitions, opening up many opportunities to discover and use this records in distinct methods for diverse purposes. The Yelp dataset contains information on local businesses in 11 cities across four continents. Based on the data, there are 86872 companies and 686556 consumers.Structure of the commercial enterprise data set for Yelp is shown in Fig. 1.

Shape of the dataset :

(10000,10)

Column Names:

Index([review_id,user_id,business_id, stars, useful, funny, cool, text, date],

dtype-'object')

Datatype of each Column:

"review_id":string

"user_id":string"

"business_id":string

"stars":int4

"useful":int3

"funny":int1

"cool":int1

"text":string

"date":string

dtype:object

Fig. 1. Structure of the Yelp

IV. Proposed Method

Fig. 2 shows the system's basic flow diagram.

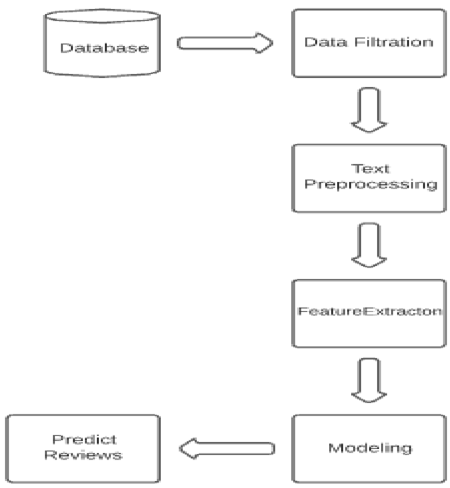

Fig. 2. Flow Diagram

i. Load Dataset

We'll start by loading the data. Every text analysis is categorized into two parts: a positive and a negative section. To start with text data and no other information, we collect them collectively.

ii. Text Preprocessing

In its raw form, any unstructured information is not well configured. Text pre-processing involves using a range of techniques to turn text with typical structure into predefined patterns. To standardize text info, we used the following per-processing strategies.

Tokenization: Split the data into letters.

Text Cleaning: Remove unwanted numbers.
Delete unwanted stopping letters like ‘at’, ’to’ etc.

iii. Data Filtration

We'll clean the data set to include 1- and 5-star ratings because their polarities are opposite, making it simpler to turn the issue into a binary classification.

iv. Feature Extraction

Since we conclude that consumer response is highly associated with how they know about the hotels, we begin the process by designing features for sentiment analysis. The feature extraction technique facilitates the translation of structured text into numerical or categorical functions. To learn, We use diverse supervised gaining knowledge of fashions on extracted features.This approach is also known as vectorization since each text is translated into a feature vector that can be fed into supervised classification systems. For feature extraction Bag of Words Model is used.

Bag-of-Words is a very sensible approach to this problem, with the following steps:

a) Implement a sequence of some samples to break the documents into tokens.
b) Each token should be assigned a weight based on how often it is present in the text.
c) Make a document-time period matrix with the aid of using putting a token in every column and a textual content in every row.

v. Predict Reviews

The satisfactory rating is going to the naive Bayes's classifier, which we use to expect a random tremendous review, a random mediocre review, and a random bad review.

V. Experiments and Results

On the raw data, we used Data Cleaning, Text pre processing, and Feature Extraction, as well as a variety of machine learning algorithms to predict user feedback based on star ratings. It assists in the animation of words as they are spoken without stuttering. Now, as seen in Fig. 3, let's print some word clouds to see what kinds of terms appear in our reviews. Birthday, Love, Concept, Place, Occasion, and other terms are all linked to the reviews. Some terms are more aligned with the customer's contact with the place.

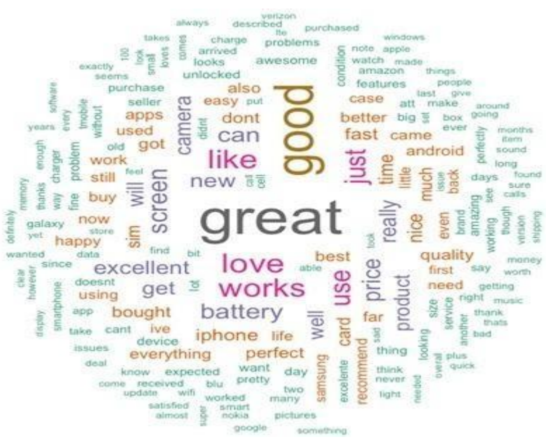

Fig. 3. User review Word Cloud

i. Data Filtration

We'll narrow down the data set to only include 1- and 5-star ratings because their polarities are opposite, making it simpler to turn the issue into a binary classification. The filtered data in Fig.4 is shown in the image below.

```
(4264, 11)
0      My wife took me here on my birthday for b......
1      I have no idea why some people give bad.......
3      Rossie, Dakota, and I LOVE Chaparral D........
4      general Manager Scott Petello is a good e.......
6      Drop what you're doing and drive here. Aft......
Name: text, dtype: object
0      5
1      5
3      5
4      5
6      5
Name:  stars, dtype: int64
```

Fig. 4. After filtration data and rating

ii.Machine Learning Algorithms

'Multinomial Naive Bayes over Gaussian is used as a classifier using the education matrix as capabilities and the textual content assessment rating (1 or 5) because the goal variable. After that, For wonderful function classification, the Naive Bayes multinomial classifier is sufficient. The multinomial feature typically desires an integer variety of features.

A random forest is a format-estimator that uses average to boost probabilistic precision and control by combining multiple decision tree classifiers on distinctive dataset sub-samples.

Decision Trees (DTs) are a non-parametrically supervised training tool for detecting and predicting patterns. The aim is to construct a system that learns from data and uses basic decision rules to predict the value of an output variable. To differentiate the different classes, the Support Vectors Classifier attempts to define the acceptable hyper plane By reducing the distance between the test points and the hyperplane, the distance between the test points and the hyperplane may be reduced.The proposed analysis employs linear SVM. Radial basis function (rbf) is the kernel type used in the algorithm. The polynomial kernel function has a degree of 'poly'. In SVM, there are 101 random state cases.

iii. The k-nearest neighbours vote classifier is used. The implemented classifier uses a total of ten neighbours. In the proposed process, uniform weights are chosen. In each neighbourhood, all points are evenly weighted. The Minkowski metric's power parameter is 2 that indicates 'euclidean distance'. The log-loss function is optimised using stochastic gradient descent in this model. In the ith secret layer, there are 100 neurons.

iv.Experiments and Results

Table I compares the accuracy, f-1, ranking, support, support, f-1 ranking and precision using different computer algorithms.

TABLE I Machine Learning Algorithms with 1 Star Rating

Machine Learning Algorithms	**1- Star Rating**		
	Precision	**Recall**	**F-1 Support**
Naive Bayes	0.85	0.75	8
Random Forest	0.79	0.39	0.53
Decision Tree	0.66	0.65	0.66
Support Vector Machine	0.75	0.8	0.77
K- Nearest Neighbors	0.75	0.1	0.18
Multi layer Perceptron	0.84	0.78	0.18

TABLE II Machine Learning Algorithms with 5 Star Rating

Machine Learning Algorithms	5 Star Rating		
	Precision	**Recall**	**F-1 Support**
Naive Bayes	0.95	0.97	0.96
Random Forest	0.88	0.98	0.93
Decision Tree	0.92	0.93	0.93
SupportVectorMachine	0.94	0.95	0.94
K- Nearest Neighbors	0.83	0.99	0.81
Multilayer Perceptron	0.95	0.93	0.93

Tables I and II show that the different machine algorithms provide superior results for 5-star rating reviews.

TABLE III Accuracy of Algorithms

Machine Learning Algorithm	Accuracy %
Naive Bayes	91.95
Random Forest	87.29
Decision Tree	87.78
Support Vector Machine	91.56
K- Nearest Neighbors	83.25
Multi layer Perceptron	93.4

Table III shows comparison with machine algorithms in respect of accuracy one-star and five-star rating reviews. The multi layer perceptron classifier achieves the highest accuracy, which is equal to 93.40 percent, as shown in the table above.

```
4    3526
5    3337
3    1461
2     927
1     749
Name:  stars,  dtype:  int64
```

Fig. 5. Prediction frequency distribution.

Figure 5 shows the predictions have a higher frequency distribution for reviews with a 5 or 4 star rating, indicating that the data set is more adjusted toward positive reviews than negative reviews.

```
My wife took me here on my birthday for breakfast and it was excellent. The weather was

Do yourself a favor and get their bloody Mary. It was phenomenal and simply the best I've

While EVERYTHING on the menu looks excellent, I had the white truffle scrambled eggs

Anyway, I can't wait to go back!
Actual Rating: 5
Predicted Rating:
5
```

Fig. 6. Actual and projected ratings for a favorable review sample.

We went here on a Saturday afternoon and this place was incredibly empty.

Next up : the wings. We were a bit hesitant to order then when the waitress

My entree was the Tilapia salad, and I was a bit disappointed. The fish was

It wasn't bad enough to say I wouldn't go back, but I won't be anxiously

Actual Rating: 1
Predicted Rating:
1

Fig. 7. Actual and projected ratings for a negative review sample

The forecast for the negative and positive review sentiment analysis is shown in Figures 6 and 7. We will learn more about the We assess that tokens of 5-star reviews are substantially reliable and 5 tokens of 1-star reviews are the most reliable by examining the most common tokens in both positive and negative feedback. Figure 8 depicts a five-token random sample for one-star and five-star rating reviews.

	one_star	five_star
Token		
treehouse	1.0	3.0
conscientious	1.0	2.0
breakfast	23.0	172.0
allowing	2.0	6.0
lax	4.0	1.0

Fig. 8 For 1 and 5 star rating ratings, the frequency of a random sample of tokens

We equate a token's score to the token's incidence by splitting it by the total no. of tokens in the set corresponding class (five star or one star). This is done so that the kindness of each token can be measured as a proxy for the kindness of other tokens in its class. The top five tokens in Fig. 8 predict one-star and five-star reviews, as well as their ratios.

TABLE IV lists the approaches used to classify emotions, along with their accuracies and data sets. As compared to other approaches, the precision of the proposed work is remarkable.

Token	**one_star**	**five_star**	**5to1_star_ratio**	**1to5_star_ratio**
staff person	0.028239	0.000375	0.013283	75.285714
refused	0.023256	0.000375	0.016129	62.000000
disgusting	0.043189	0.000750	0.017370	57.571429
filthy	0.018272	0.000375	0.020528	48.714286
unacceptable	0.016611	0.000375	0.022581	44.285714

TABLE V displays the methods and data sets used to categorize attitudes, along with their accuracies.

Sr. No.	Data set	Method Used	Accuracy %
1	Manually classified tweets	Gradient Boosting[18]	81.82
2	Twitter API	Random Forest[3]	81.69
3	Twitter API	Linear Regression[4]	85.23
4	Real Chinese Restaurant Reviews	Multi Aspect Bootstraping[8]	75.5
5	Yelp Dataset	SVM[19]	88.9
6	Yelp Dataset	Naive bayes[10]	92.6
7	Yelp Dataset	Bagging MLP[20]	92.1
8	Yelp Dataset	Naive Bayes MLP [roposed]	93.15 93.4

VI. Conclusion

In this paper, We identified an investigation into determining the polarisation of a user's textual analysis as positive or negative automatically. There is a strong need for such research because rankings and stars are becoming increasingly relevant in assisting prospective clients in making decisions or purchasing products. For our tests, we used the Yelp data collection. To predict the review, we used a machine learning algorithm. Using MLP, our system achieved a recognition rate of 93.4 percent, which is higher than the current system's recognition rate.

REFERENCES

[1] Asghar, N.,"Yelp dataset challenge: review rating prediction."

[2] Thakkar, Harsh & Patel, Dhiren. (2015). Approaches for Sentiment Analysis on Twitter: A State-of-Art study, https://arxiv.org/abs/1512.01043

[3] Arti, K. P. Dubey and S. Agrawal, "An Opinion Mining for Indian Premier League Using Machine Learning Techniques," 2019 4th International Conf. on IoT: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-4.doi: 10.1109/IoT-SIU.2019.8777472

[4] A. Razia Sultana, A. K. Jaithunbi, and L. Sai Ramesh, "Sentiment analysis in twitter data using data analytic techniques for predictive modeling", J. Phys. Conf. Ser., vol. 1000, no. 1, 2018 pp. 1-4.

[5] L. Kao and Y. Huang, "Predicting purchase intention according to fan page users' sentiment," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 831-835, doi: 10.1109/SMC.2017.8122712

[6] F. Ren and Y. Wu, "Predicting User-Topic Opinions in Twitter with Social and Topical Context," in IEEE Transactions on Affective Computing, vol. 4, no. 4, pp. 412-424, Oct.-Dec. 2013, doi: 10.1109/T-AFFC.2013.22

[7] R. Chakraborty, M. Bhavsar, S. K. Dandapat and J. Chandra, "Tweet Summarization of News Articles: An Objective Ordering-Based Perspective," in IEEE Transactions on Computational Social Systems, vol. 6, no. 4, pp. 761- 777, Aug. 2019, doi: 10.1109/TCSS.2019.2926144

[8] J. Zhu, H. Wang, M. Zhu, B. K. Tsou and M. Ma, "Aspect-Based Opinion Polling from Customer Reviews," in IEEE Transactions on Affective Computing, vol. 2, no. 1, pp. 37-49, Jan.-June 2011.doi: 10.1109/TAFFC.2011.2

[9] B. S. S. Govind, D. R. Tene and K. L. Saideep, "Novel Recommender Systems Using Personalized Sentiment Mining," 2018 IEEE International Conf. on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, 2018, pp. 1-5, doi: 10.1109/CONECCT.2018.8482394

[10] J. Wang, "Predicting Yelp Star Ratings Based on Text Analysis of User Reviews," 2015

[11] J. Chen et al., "Double layered recommendation algorithm based on fast density clustering: Case study on Yelp social networks dataset," 2017 International Workshop on Complex Systems and Networks (IWCSN), Doha, 2017, pp. 242-252,doi: 10.1109/IWCSN.2017.8276534

[12] Yelp dataset challenge, http://www.yelp.com/dataset_challenge

[13] B. K. Bhavitha, A. P. Rodrigues and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentiment analysis," 2017 International Conf. on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2017, pp. 216-221, doi: 10.1109/ICICCT.2017.7975191

[14] R. B. Shamantha, S. M. Shetty and P. Rai, "Sentiment Analysis Using Machine Learning Classifiers: Evaluation of Performance," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 21-25, doi: 10.1109/CCOMS.2019.8821650

[15] http://machinelearningmastery.com/

[16] H. Valecha, A. Varma, I. Khare, A. Sachdeva and M. Goyal, "Prediction of Consumer Behaviour using Random Forest Algorithm," 2018 5th IEEE Uttar Pradesh Section International Conf. on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, 2018, pp. 1- 6,doi:10.1109/UPCON.2018.8597070

[17] J. Shi, X. Zhu and N. Niu, "A Research on Multi-Layer Perceptron Diagnosis Model Based on D Matrix," 2018 Prognostics and System Health Management Conference (PHM-Chongqing), Chongqing, 2018, pp. 769- 773,doi: 10.1109/PHM-Chongqing.2018.00137

[18] A. G. Prasad, S. Sanjana, S. M. Bhat and B. S. Harish, "Sentiment analysis for sarcasm detection on streaming short text data," 2017 2nd International Conf. on Knowledge Engineering and Applications (ICKEA), London, 2017, pp. 1-5. doi: 10.1109/ICKEA.2017.8169892

[19] Yu Boya, Zhou Jiaxu, Zhang Yi, Cao Yunong, "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews", ARXIV, 2017.

[20] G. S. Budhi, R. Chiong, I. Pranata and Z. Hu, "Predicting rating polarity through automatic classification of review texts," 2017 IEEE Conference on Big Data (ICBDA), Kuching, 2017, pp. 19-24. doi: 10.1109/ICBDAA.2017.8284101

[21] Shahida Khan, Kamlesh Chopra, Pratyush Sharma, “Brand Review Prediction using User Sentiments: Machine Learning Algorithm”.