International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 10, October 2022)

Use of Ml Algorithms for Automatic Sentiment Analysis of User Reviews

Shubham Kumar[1], Dr. Pankaj Richhariya[2]
Bhopal Institute of Technology and Science, Bhopal (M.P.), India

***Abstract*-- Everyone is sharing their stories on social networking platforms in increasing numbers. Every commercial website online heavily relies on customer feedback. Shops, malls, and a variety of other business kinds can now be found online. They merely want customers to place them among the top-rated businesses online. Customers' personal use of the product had an impact on how they perceived the service. A "as info" symbol on the internet represents user opinions. A research project on hotel mood has been proposed. Customers are able to rate and contrast our hotel using this approach.**

***Index Terms*--Machine Learning, Naive Bayes, Natural Language Processing, Recommender Systems, Sentiment Analysis.**

I. INTRODUCTION

Currently, there are businesses or organisations whose success is based on client feedback. The new company will require more operators because it will offer more items and services. It follows that they must care about the opinions and ratings of their users if they want the organisation to grow. On online portals, clients have the option to voice their ideas. We can estimate the company's potential rating based on these details. To get extremely precise and finely grained evaluations, it is important to employ a system of one or more suggestions. When it comes to ratings, visitors give hotels a lot of weight to factors like maintenance, customer service, etc. Customers' opinions of hotels are significantly influenced by the services they receive, including free internet, the hotel's location, and amenities like a bar lounge, wheelchair accessibility, childcare services, and babysitting rooms. if you only decide. This one does a better job of expressing the sentiment than that one, while the latter is horrible. Customers typically wish to give their feelings in this situation more weight by adding ratings and reviews.

One item or a series of things may receive recommendations from a recommender system. In most cases, e-commerce websites employ a single item to determine the sale and product of a single item, while a sequence recommender system for a sequence of items is used to predict the business of any organisation, including the hotel industry, cleanliness, services, and food. A good recommender system is one that can recognise the right users, their motivations, expectations, and goals.

The recommender system is capable of effectively comprehending the item set's characteristics, including volume, distribution, nature, and user ratings.

To be deemed a powerful recommender system, the recommendation technology should apply "trust metrics." Research has shown that people prefer recommendations from people they trust, such as their friends, over those from internet recommender systems. The trust enhanced recommender system must use information that comes from trusted networks, such as friends, people who share your interests, people who live in the same city, etc. We must apply group recommendations to specific users in order to create a powerful and trustworthy recommender system. Three uses for group recommendation techniques are as follows: (1) to combine several criteria, (2) to address the so-called cold-start issue, and (3) to consider other people's perspectives.

II. LITERATURE SURVEY

[1] Nibedita Panigrahi and Asha T. devised a method in 2018 that employed RHALSA (Ranking Hotels Using Aspect Level Sentiment Analysis) algorithm to rate hotels. The information was gathered from a few TripAdvisor reviews. The two main components of the proposed work were cleaning and hotel service. The proposed method for sentiment analysis to calculate scores employed the Standard Core-NLP sentiment approach. The Stanford Core-NLP Sentiment Levels are used to produce the Sentiment Score, which is displayed as 0 for "Very Negative," 1 for "Negative," 2 for "Neutral," 3 for "Positive," and 4 for "Very Positive." The outcome demonstrates that there are divergent views regarding each factor that was taken into consideration for the same hotel. The RHALSA that was given could only handle disparaging remarks, but it might be expanded to include discourse relations that might alter the sentence's orientation.

[2] The Support Vector Machine algorithm was used by the authors of this research to develop a sentiment classifier. A machine learning system is trained to determine the frequency of particular words using the Unigram language model. The TF-IDF was used as a strategy and to track down instances of particular polarity terms.

The dataset used for the tenfold cross validation was divided into ten equal-sized folds with 180 reviews each. At that time, the cross-validation procedure was used. Extremely brief reviews—those with less than 30 words—and extremely long reviews—those with more than 250 words—are excluded from the corpus in the proposed approach. The accuracy of the TF-IDF was 95.78%, and that of the TO-approach was 71.76%, according to the authors.

[3] People provide reviews and assign a numerical score to hotels based on their perceptions. There are numerous factors at play in their reviews. Most people glance at the hotel's overall rating but do not read every review. Sentiment analysis that is particular to an aspect so offers a viable answer. On this topic, a work using the ILDA (Interdependent Latent Dirichlet Allocation) algorithm was provided. The suggested approach divides the review vocabulary into header and modifier terms. For instance, if a review states, "Lovely service," then "nice" is the modifier term and "service" is the header term. There is no sentiment polarity in the header words. The sentiment polarity is caused by the modifier phrases. The header word cannot be modified, however the modifier terms can be changed by using terms like lovely, worst, good, and so on. The header phrase provides the aspects, whereas the modifier term provides the polarity.

ARIH is the algorithm employed (Aspect and Rating Inference Using Hotel Specific Aspect Rating Priors).

[4] Formally, we typically define a corpus of information made up of examined N documents as D=x1,x2,xD. Each document that is inspected xD is made up of a 26-token sequence (WWW(2017) 20:23–37). Each review xd has a corresponding overall rating rd, which ranges from 1 to S (S = 5) and takes the form of an integer. An attribute of a hotel that is predetermined, such as price, room, location, and service, is called an aspect. A text review expresses the reviewers' opinions based on several factors. The use of the word "price" serves as an example of how the review comments on aspect value are communicated. Each review has a number of integer scores denoted by the letters l1, l2, and lK, where K is the total number of aspects. Phrase We presum that every review is a collection of opinions.

[5] The author of this article classified sentences as subtrees of dependency trees. classifying the polarity of a sentence based on sub-opinions in a dependency tree using many factors, such as word granularity, sum-product belief propagation, and sub-opinion relations. This study employs dependency sub-tree-based sentiment categorization algorithm.

The dataset utilised has an id and a comment's overall score. For analysis, a further sentence is tokenized and processed. Recall and precision calculations will therefore produce the findings.

[6] One research suggested sentiment classification using two different methods. The first is a machine learning approach, whereas the second is a semantic orientation approach. Bag of Words is a preset dataset that is utilised in two stages of training and prediction in the machine learning approach. Finding the key words that best describe the sentiment words in the text is the focus of the semantic orientation technique. Finding the sentiment of a text or sentence can be done using both machine learning and a semantic method.

[7] One method used in research papers [6] was topic modelling and text mining. An approach to text mining that is widely used to uncover hidden semantic patterns in text bodies is topic modelling. The topic modelling method is used in this paper to determine word frequency. The writers I suggest topic modelling as a potential future research avenue to investigate for sentiment classification.

[8] One of the approaches suggested in the research was classification using machine learning techniques. Opinion mining is a technique for analysing texts and extracting information. Two machine learning classification techniques, such as Nave Bayes and Decision trees, are used for this. Even still, Nave Bayes is a more trustworthy classification method than decision tree since choice tree may not produce as accurate findings as Nave Bayes when the size of the data increases. Data sets are split into training and testing sets for machine learning algorithms so that additional analysis on these sets might yield results.

[9] One of the research papers provided offers sentiment classification in two different ways. For determining the orientation of a document of words or phrases, the first approach is lexicon-based. The second approach uses machine learning techniques like clustering. Applying TF-IDF for this purpose provides the term frequency of terms in a document. By assessing sentiment strength, we can obtain specific values and use clustering to apply those values. We categorise the words/text into positive, negative, and neutral by computing Euclidean distance.

[10] The authors of this paper employed a technique known as contradiction detection. Through data analysis and pre-processing, the contradiction detection is located. They did this by converting dates and time formats into a range of t and h, which they used to demonstrate how numerical mismatch causes disagreement.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 10, October 2022)

A text is in contradiction if it does not fall between t and h. The suggested contradiction focuses on feature selection and data pre-processing for accuracy.

III. Proposed System

In the proposed work we are using machine learning algorithms for sentiment analysis for hotel business. The machine learning techniques can do work on large amount of data sets. The real-world data can contain noisy values in it, for that some cleaning process is required which is known as pre-processing step. That reduces undesirable data impact present in information which augmenting its data. The real- world data always requires to be clean and transform inorder to be usedby machine learning techniques. The pre- processing step includes sampling also to reduce big population intoprecise number of data-set. Somecommon pre-processing techniques are collaborative filtering, sampling, reducing dimensionalities, principal component analysis, singular value decomposition etc.,

A. Machine Learning

The goal of Machine Learning is to grasp the structure of the data and convert that knowledge into models which will be understood and utilized. Machine learning is the sub branch of artificial intelligence and the advance concept of machine learning is deep learning. Machine learning will be able to predict the future based on the past or historical data. The most commonly used Machine Learning algorithms are decision tree, k-means clustering, support vector machine, random forest, neural network.

B. Classification

In Machine Learning and statistics, classification is the problem of identifying to which of a set of categories (sub- populations) a new observation belongs, on the basis of a trainingset of data containing user reviews. Examples are assigning a rating to the hotel based on positive and negative reviews. Classification is an example of patternrecognition. Classification comes under supervised learning, which learns a function and maps an input to the output. some of the classification algorithms are Naïve Bayes, decision treesetc.,

IV. Methodology

A. Naive Bayes:

Bayes theorem provides a way to calculate the probability of a hypothesis based on its prior probability and the probabilities of observing various data.

$$P(h/D) = (P(D/h) * P(h) \qquad (1)$$

- A concept learning algorithm considers a finite hypothesis space „h" defined over an instance space X.
- What can we do if our data *d* has several attributes?
- Naïve Bayes assumption: - Attributes that describe data instances are conditionally independent given the classification hypothesis.

$$P(d|h) = P(a1,\ldots,aT|h) = \coprod t\ P(at|h) \qquad (2)$$

- It is a simplifying assumption, obviously it may be violated in reality
- In spite of that, it works well in practice
- The Bayesian classifier that uses the Naïve Bayes assumption and computes the MAP hypothesis is called Naïve Bayes classifier
- One of the most practical learning methods

B. Natural Language Processing (NLP):

Natural Language Processing, or NLP for short, is extensively characterized as the programmed control of natural language, similar to speech and text, by programming. NLP is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to method and analyse massive amounts of language information. It is the branch of machine learning which is about analyzing any text and handling predictive analysis.

NLP is useful in hotel rating for dividing the sentences and determines whether the sentence is Positive, Negative and Neutral and also, NLP will also be useful as translator in case translation of one language to required language.

V. Experimental Results

In this work we implemented machine learning algorithms On the data collected from various websites such as trip advisor, trivago etc., The Naïve Bayes classifier is used to classify the sentiments into two broad categories i.e., positive and negative. The dataset contains sentimental statements from various reviews.

Steps:

1. Create a dictionary that has the data with labels(positive/negative).
2. Create the dictionary, feature set = { }, that contains the count of occurrences of word undereach label.
3. Split the sentences with respect to non-characters and key words.
4. Use Naïve Bayes classifier with respect to non-characters and key words.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 11, Issue 10, October 2022)

Weuse 60% data as training set and 40% data as testing set. The classifier was perfectly able to find out positive and negative classes on the basis of lexicon values. In lexicon we used 45 key values for positive and 45 for negative.

Figure 1 Naive Bayes results

In the figure 1 shows the classification of the reviews as positive or negative. For this the input takes five sentences and classify each of the review. Finally it shows the total number of positive ratings and negative ratings.

Figure 2 NLP result 1

For figure 2 input is a single sentence and it calculates the overall sentiment dictionary score for the positive, negative, and neutral. It calculates the rating of positive, negative, neutral with the help of compound rate. Finally it gives the review overall rating as positive or negative or neutral.

Figure 3 NLP result 2

As mentioned in figure 2 here in figure 3 we have given different reviews as input and calculated the overall review rating.

VI. Conclusion

The content-based recommender implies matching of attributes from a user profile in which preferences and interest are stored with attributes of content object. If a string, or some morphological variant, is found in both the profile and the document, a match is made and the document is considered as relevant. The major limitation is matching the keywords and the overall vocabulary. The Naïve Bayes classification algorithm is good for scaling the dataset and implement the linear equation on features and predicators. The classifier got some mismatched values for neutral results.

REFERENCES

[1] Nibedita Panigrahi and Asha T, ”RHALSA: Ranking Hotels using Aspect Level Sentiment Analysis”,

[2] George Markopoulos, George K. Mikros, Anastasia Iliadi “Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features Article”, pg:-373-383

[3] by Wei Xue, Tao Li, Naphtali Rishe, “Aspect identification and ratings inference for hotel reviews”, pg:-24-37

[4] Tran Sy BANG and Virach SORNLERTLAMVANICH, “Sentiment Classification for Hotel Booking Review Based on Sentence Dependency Structure and Sub-Opinion Analysis”, pg:-910-916

[5] Youngseok Choi & Habin Lee, “Data properties and the performance of sentiment classification for electronic commerce applications”,pg:- 994-1012

[6] Sentiment classification of consumer generated online reviews using topic modeling

[7] Wararat Songpan, “The Analysis and Prediction ofCustomer Review Rating Using Opinion Mining”, pg:-71-77

[8] Sumbal Riaz, Mehvish Fatima, M. Kamran, M. Wasif Nisar, “Opinion mining on large scale data using sentiment analysis and k-means clustering”

[9] Siti Nuradilah Azman et al, “Towards an Enhanced Aspect-based Contradiction Detection Approach for Online Review Content”

[10] Francesco Ricci · Lior Rokach · Bracha Shapira · Paul B. Kantor(Editors), Recommender systems handbook, springer, 2011.