AIRO JOURNALS

Free / Unpaid
Peer Reviewed
Multidisciplinary
National

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

APPLICATION OF MACHINE LEARNING TO ANALYZE REVIEWS OF MOVIES, CUSTOMER COMMENTS, AND SUPPORT TICKETS

Sagar Dixit
Research Scholar

Abstract

Nowadays, computerized evaluations have a big impact on consumer purchasing behavior and cross-border customer communications. Customers are given a forum by internet retail giants like Amazon, Flipkart, and others to educate and inform potential purchasers about a product's presentation. To distinguish meaningful experiences from a vast collection of evaluations, it is vital to divide evaluations into positive and negative sensations. Feeling investigation is the process of examining one's thoughts and conclusions after reading any form of material. Sentiment analysis is also known as assessment mining. The viewpoint of the mass, the group, or any person can be provided by feeling analysis of the data. This technique is employed to ascertain a person's viewpoint regarding a specific material source. There is a substantial amount of material available for entertainment on websites, tweets, blogs, status updates, and other web-based platforms.

Keywords: *Machine Learning Implementation, Movie Reviews, Customer Feedback, Support Tickets*

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

I. INTRODUCTION

The growth of digital innovation has significantly altered how people communicate their thoughts. When people shop online or buy movie tickets to see films in theaters, they rely on this user-generated content to evaluate the products. The clients interact with one another through postings, Facebook, tweets, hash tags, and other social media. Since there is so much material available, it would be challenging for a typical individual to study and comprehend it all. The primary issues with sensation analysis are the arrangement of presumptions and discernible evidence. It is clearly explained as falling into two categories: information-based approaches and machine learning techniques. [1] To recognize sentiments, the first method requires a sizable library of prepared emotions and an expert information depiction. However, the machine learning approach uses an informative collection connected to tests and preparation to construct a classifier. Information base methods with less complexity are chosen. The development of calculations led to the discovery of a few challenges in the area of feeling study. The first is that an assessment phrase may be favorable or negative depending on the circumstances. The next test is that opinions are rarely expressed by different persons in the same manner. Assessment mining facilitates the understanding of the relationship between text-based reviews and their results.

One of the major forces influencing human behavior is emotion. We routinely rely on or steal ideas from others in practically every facet of life, even clearly demonstrated distraction. Because to the extensive use of the internet, it is rather typical for people to post reviews of their ideas, theories, and insights in a range of open forums and private online journals. These studies or situations with fundamentally nonexclusive handling must be finished in order to provide precise conclusions. Opinion examination is important in getting input from consumers on both freshly delivered items and convenience issues with already-existing items. As this type of feedback component grows popularity, an increasing variety of industries are adopting it, including those that demand temperamental content, such as song verses, news comments, sports and games, movie reviews, etc. Everyone must follow it, including various government agencies and business projects. These associations greatly benefit from the assistance of Opinion Evaluation while assessing the presentation of their goods.

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

The best type of entertainment that anyone can enjoy is unquestionably watching movies, and it is typical for people to watch movies and then debate them online in public forums or private web journals. Surprisingly, the viewers are also significantly impacted by these assessments, in addition to the filmmakers. Instead of reading the vast amount of content that consumers have provided, we can analyze the text-based data using opinion research to assess the overall impact that movies have had on people.

The ongoing demand for cutting-edge cells is what is driving the mobile phone industry's rapid expansion. With the PDA market's rapid expansion, it is crucial to properly investigate and comprehend brands and phone models. There are numerous brands to consider, some of which have a sizable and significant market share. I do [2] have some. For instance, words like Samsung and Apple are linked to well-known worldwide businesses. E-commerce has a big impact on consumer purchasing behavior, which is why sales of mobile phones are increasing. Consumers can use the reviews that are published on these online marketplaces as a guide to make knowledgeable judgments. With the wide variety of goods manufactured by different companies, it is critical to give consumers accurate ratings. The number of reviews of a product or brand is growing at a concerning rate, much like the large volume of information that needs to be managed. Providing the survey's feeling direction by classifying reviews based on user opinions into positive and negative sentiments improves judgment. Reviews that are divided into opinions may make it easier for potential customers to assess both positive and negative feedback and decide what is best for them. This evaluation acts as a statement to the customers who are interested in learning about the details and features of the cell phones, fostering greater customer confidence.

II. LITERATURE REVIEW

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

Many various degrees of detail, such as the report level, the phrase level, and many more, have been addressed in relation to opinion investigation. In this section, we provide you with some of the pertinent research that has previously transformed this area of study.

According to [3], using a corpus of 1131 positive and negative test data, the Guileless-Bayes classifier surpasses the other classifiers, Guileless-Bayes, Perceptron, and Rocchio classifiers, in terms of accuracy. . In relation to language learning, Facebook data was gathered.

In [4], the author used ID3, Credulous Bayes, and Support Vector Machines (SVM) in 175 cases to show how his TeachingSenti lexicon was created to demonstrate the directed definitional evaluation of mechanized limb opinion. With an estimated accuracy of over 97%, the SVM classifier fared better than the other classifiers.

During the "Arab Spring," the author of [5] compared the effects of her SVM classifier and her Credulous Bayes classifier on her Facebook status of Tunisian clients.

The Opinion Assessment technique utilized by the authors in [6] at different levels, specifically local feeling that adds up to global feeling, outperforms the general SVM classifier.

In order to cognitively characterize 185 melodic verses, the writers of [7] employed Senti Word Net to find significant terms. They also used SVM, K-Nearest Neighbor (KNN), Sincere Bayes Classifier, and Artless Bayes Classifier. In comparison to other estimates, this is better.

For the purpose of evaluating the cycle, Rawan T. Khasawneh obtained data using Senti Strength and his Social Mention Contraption in [6]. In [8], JalelAkaichi and his colleagues used bigram, trigram, and unigram collocation highlights to examine the accuracy of SVM and his Sincere Bayes Classifier. AddlightMukwazvure and K.P. used SVM and ANN classifiers for assessment mining. Supreethi in shown how SVM outperforms ANN in a number of ways. K used a collection of word models that combined SVM classifiers for handling common languages, linguistic names, and feathery request math. Mouthami and the others.

Sudipto Shankar Dasgupta and his associates examined monitoring of several brands using CRAN's Hadoop Guide Minimizer and R Feeling packages, and in [9] he determined the accuracy

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

of five different brands. In [11], the author uses his HDFS structure in Apache Hadoop to create a high-speed information transmission mechanism and employs a hybrid structure that makes use of both production hardware and open source software to create IT initiatives. improved the effectiveness of

Information analysis has made it possible to reveal information's hidden examples. Big Information is characterized by volume, speed, and variety [10]. Two other factors that play a key role in the vast majority of information are veracity and worth. The amount and rate at which new information is being created on a constant basis exceeds the capacity of many IT divisions. Websites for online businesses are crammed with a huge assortment of various reviews for various products. These evaluations can be used to understand consumer behavior and guide choices. Reviews might be organized or unorganized. Important business knowledge can be gathered by separating the relevant data from the irrelevant data. Large-scale information has allowed businesses to thrive and rely more on logic than intuition.

III. MACHINE LEARNING METHODS

A. Naïve Bayes

It is a process in light of the Bayes hypothesis. According to the Guileless Bayes classifier, the presence of one element in a class will not have an impact on the presence of another. For extremely huge datasets, this model is incredibly helpful and easy to construct. In addition to being simple, gullible Bayes is known to outperform even more advanced order algorithms.

$$\boldsymbol{P\,C|X\ =\ P(X|C)\ *\ P(C)/P(X)}$$

P(C|X) is posterior probability of class C

P(C) is prior probability of class C

P(X|C) is probability of predictor given the class.

P(X) is prior probability of predictor

B. K- Nearest Neighbour

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

The simplest machine learning calculation is K-NN. This method's main idea is to select a particular number of training prerequisites that are actually closest to the new point and predict a name from it. Depending on how evenly distributed the central districts are, the number of tests may either remain constant or fluctuate. Any capability of distance estimation is possible. [11] The most used formula for calculating the separation between two focuses is the standard Euclidean distance. Many design and recurring difficulties, such as managing translated digits and satellite symbolism information, have been handled using Closest Neighbors.

C. Random Forest

The teaching technique for characterization and relapse is arbitrary woodlands. During preparation, various option trees are built. It sends the new case to each tree to arrange it. Each tree follows a particular arrangement, and the outcome is a class. It is believed that irregular woodlands produce the most extreme number of comparative class produced by varied trees, the class decided by majority vote.

Both experts and non-specialists can readily grasp and use arbitrary woodlands with little study and programming. It is easily usable by those without extensive backgrounds in their areas of expertise.

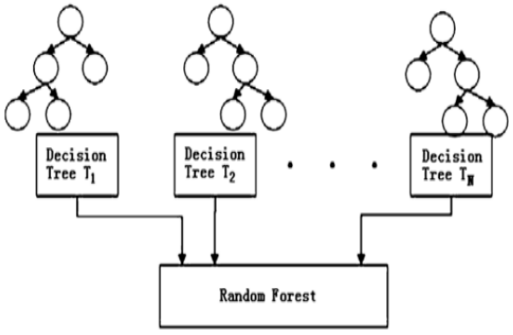

Figure: 1. a fictitious forest figure from Techleer.

IV. EXPERIMENTAL SETUP

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

An outline was created for each paper by reading and examining it before the inquiry began by reviewing several examination and audit papers on sensory investigation. Support Vector Machine, Choice Tree Enlistment, K-Nearest Neighbor, Innocent Bayes, and Irregular Woodland were investigated as well as other frequently used arrangement computations. The arff design was subsequently made using this knowledge [12]. The information was contained in a record with the name "txt token," which had two suborganizers: one for positive and one for negative data. The new information was added to the WEKA system using Text Index Loader. Text pre-handling was then done on the WEKA system.

Open source software known as WEKA is made available under the GNU General Public License. The framework was created at the College of Waikato in New Zealand. The software is called WEKA, which stands for Waikato Climate for Knowledge Examination. It offers an implementation of machine learning calculations. It has information handling modules.

Information preprocessing, binning, grouping, relapse detection, and component identification are a few information mining activities that WEKA enables. WEKA lacks multi-social information mining capabilities, however there is distinct programming for combining many connected data set tables into a single table that is utilized for processing. As part of the characterisation problem, a directed learning job, a class name must be assigned to an unclassified tuple in line with a set of generally grouped occurrences that serve as a practice set for the calculation.

Credulous Bayes, k-nearest neighbors (k = 3), and Arbitrary Backwoods are our three characterization techniques. The number of cases that are correctly arranged will be used as a quality indicator. [13] We used 10 overlay cross approval during the approval stage. The methods used to conduct testing are next:

Step 1: Putting the dataset into WEKA is stage one. Importing the dataset into the WEKA apparatus was the fundamental action taken. This stage was performed using a basic import method for printed datasets called Text Library Loader component.

Step 2: After being imported, the dataset is converted and saved in the ARFF format.

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

Step 3: Next, 2000 cases are joined with the credits "Text" and "Class" to create a relationship. The uniform circulation of the property Class is shown in the illustration. As shown in Fig. 2. In terms of evaluations of negative extremity, the blue tone is used, whereas the red tone is used for evaluations of positive extremity.

In Step 4, the String To Word Vector filter is then applied.

Next, in Step 5, the Attribute Selection filter is applied.

Step 6: After applying the Attribute Selection filter, the outcomes are as shown in Fig 3.

Step 7: Using the information produced by the aforementioned developments, three calculations are performed. The three calculations are Credulous Bayes, K Closest Neighbor, and Arbitrary Backwoods.

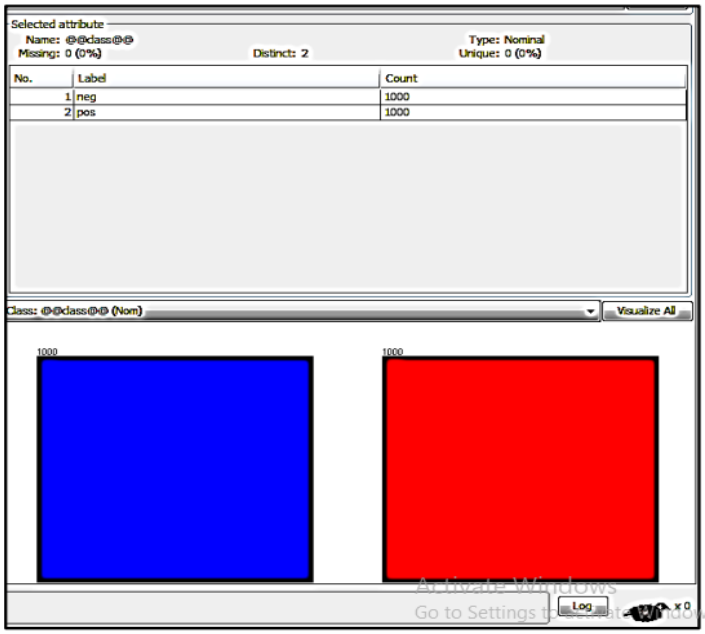

Figure: 2. after loading the data in aRF format, the Weka interface

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

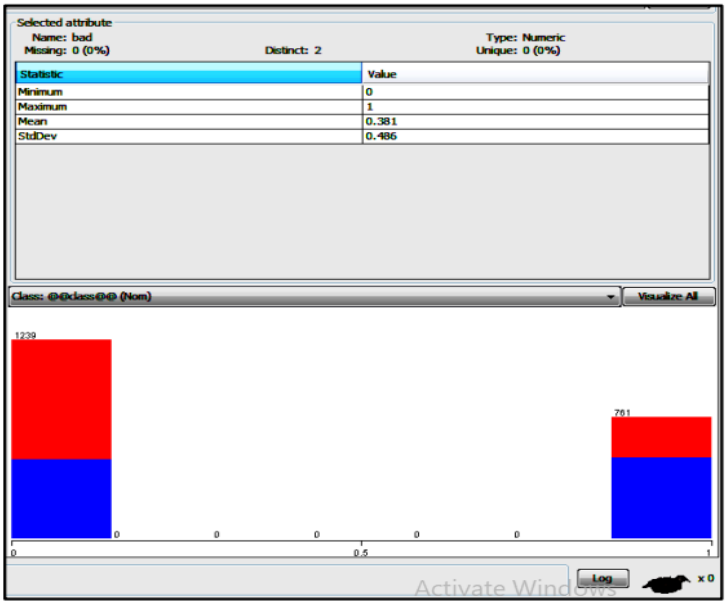

Figure: 3. After Attribute Selection Filter, the Weka interface.

V. RESULTS

Navie Bayes, K-Nearest Neighbor, and irregular backwoods calculations Calculations were done using the information that was gathered. [14] The results are summarized in the tables and figures that are supplied below. The calculation using Guileless Bayes had the highest level of precision.

Table: 1. Accuracy Ratios for Different Algorithms

Naïve Bayes	K Nearest Neighbour	Random Forest
72.3%	44.20%	87.56%

Table: 2. Different Algorithms' Mean Absolute Error

Naïve Bayes	K Nearest Neighbour	Random Forest
0.3223	0.3364	0.2387

Table: 3. Different Algorithms' Mean Absolute Error

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
National
ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

Naïve Bayes	K Nearest Neighbour	Random Forest
0.2682	0.5663	0.2814

VI. CONCLUSION

the comparative analysis of different machine learning formulas that can be used to extract emotions from text. [15] They are more productive and simple. With its extreme precision, Innocent Bayes and SVM can be used as a standard for many other types of calculations. It provides enough information to help with further research and working on the forecasts. The presentation of a formula that predicts the rate of movie production is expected to benefit from cleaner data. Several methods were employed to distinguish the extremes of the tweets. The calculations made use of Credulous Bayes, K-Nearest Neighbor, and Irregular Backwoods. The most effective classifier was Credulous Bayes. There hasn't been much experimentation with the calculations, therefore it is expected that new estimations will be tested or that a cross-approach will be developed to improve the accuracy of the results. As a general rule, determining a survey's limits can be useful. You can develop an intelligent system that can provide customers with comprehensive reviews of movies, products, services, and other media without the client reading individual reviews. In light of everything, decisions can be made simply based on the outcomes of sophisticated frameworks.

VII. FUTURE SCOPE

Studies on sensation evaluation have been highly active recently due to a number of testing research problems and practical applications. Although this inquiry has filled a few holes, it is predicted that future efforts will also result in improvements. The most necessary change for this inquiry is the enlargement of the opinion dictionary because there aren't many terms for feelings in the current lexicon. It is also required to create various feeling dictionaries for different contexts, like a movie theater. In numerous ways, the proposed framework can be put to the test and further assessed. Future research may also incorporate methods for creating dictionaries, which can save time and lessen the need for human labor and expand the scope of this investigation.

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

In this study, Sentlex, another framework for analyzing opinions that also uses a semantic direction approach, has been compared to the presentation of the suggested framework Clari Sent. A machine learning strategy can be used to perform execution correlation between Clari Sent and many frameworks.

REFERENCES

1. *AddlightMukwazvure, K.P Supreethi,"A Hybrid Approach to Sentiment Analysis of News Comments, 2015" ,Reliability, Infocom Technologies and Optimization(ICRITO), 2-4 Sept. 2015.*
2. *Charit Pong-inwong, WararatSongpan ,"TeachingSentiLexicon for Automated Sentiment Polarity Definition in Teaching Evaluation", Proceedings of Semantics, Knowledge and Grids (SKG), 2014 10th International Conference on 27-29 August 2014.*
3. *Christos Troussas, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno, Jaime Caro, "Sentiment Analysis of Facebook statuses using Naïve Bayes classifier for language learning", Proceedings ofInformation, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on 10-12 july 2013*
4. *D. N. Devi, C. K. Kumar, and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in Advanced Computing (IACC), 2016 IEEE 6th Interna-tional Conference on. IEEE, 2016, pp. 3–8.*
5. *Gaurav D Rajurkar, Rajeshwari M Goudar, "A speedy data uploading approach for Twitter Trend And Sentiment Analysis using Hadoop", Proceedings 2015 International Conference on Computing Communication Control and Automation.*
6. *JalelAkaichi, "Social Networks „ Facebook" Statuses Updates Mining for Sentiment Classification", proceedings of SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*
7. *JalelAkaichi, zeinebDhouioui, Maria Jose LopezHuertasPerez,"Text Mining Facebook Status Updates for Sentiment Classification, 2013", Proceedings of System Theory, Control and Computing(ICSTCC), 2013 17th International Conference*

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

8. *K. Mouthami, K. Nirmala Devi, V. Murali Bhaskaran," Sentiment Analysis and Classification Based On Textual Reviews, 2014", Proceedings of Information Communication and Embedded Systems (ICICIES), 21- 23 Feb 2013.*
9. *M. WAHYUDI and D. A. KRISTIYANTI, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." Journal of Theoretical & Applied Information Technology, vol. 91, no. 1, 2016.*
10. *Na Fan, Wandong Cai, Yu Zhao,"Research on the Model of Multiple Levels for Determining Sentiment of Text", Proceedings of 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application.*
11. *P. Russom et al., "Big data analytics," TDWI best practices report, fourth quarter, pp. 1–35, 2011*
12. *Rawan T. Khasawneh, Heider A. Wahsheh, Mohammed N. Al-Kabi, Izzat M. Alsmadi, "Sentiment Analysis of Arabic Social Media Content: A Comparative Study, 2013", Proceedings of the 8th International Conference for Internet Technology and Secured Transactions(ICITST2013)*
13. *S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," Journal of Business Research, vol. 69, no. 2, pp. 897–904, 2016.*
14. *Sudipto Shankar Dasgupta, Swaminathan Natarajan, Kiran Kumar Kaipa, Sujay Kumar Bhattacherjee, Arun Viswanathan,"Sentiment Analysis of Facebook Data using Hadoop based Open Source Technologies, 2015" Proceedings of Data Science and Advanced Analytics(DSAA),2015 19-21Oct.2015.*
15. *Vipin Kumar, SonajhariaMinz,"Mood Classification of Lyrics using SentiWordNet" ,Proceedings of 2013 International Conference on Computer Communicationand Informatics(ICCCI-2013), Jan. 04-06, 2013, Coimbatore, INDIA*

Author's Declaration

I as an author of the above research paper/article, hereby, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I

ISSN: 2321-3914
Volume 1 Issue 3
March 2023
Impact Factor: 10.2
Subject Computer Science

shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website/amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct I shall always be legally responsible. With my whole responsibility legally and formally I have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and the entire content is genuinely mine. If any issue arise related to Plagiarism / Guide Name / Educational Qualification /Designation/Address of my university/college/institution/ Structure or Formatting/ Resubmission / Submission /Copyright / Patent/ Submission for any higher degree or Job/ Primary Data/ Secondary Data Issues, I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the data base due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Aadhar/Driving License/Any Identity Proof and Address Proof and Photo) in spite of demand from the publisher then my paper may be rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds any complication or error or anything hidden or implemented otherwise, my paper may be removed from the website or the watermark of remark/actuality may be mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me

Sagar Dixit