e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711050

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Comparative Study of Machine Learning Approaches for Airline Reviews

Santhosh Ram.K[1], Akshatha.R[2], TinckyShalinee.A.S[3]

Student, Department of Software System, Sri Krishna Arts and Science College, India[1]

Assistant Professor, Department of Software System, Sri Krishna Arts and Science College, India[2]

Assistant Professor, Department of Software System, Sri Krishna Arts and Science College, India[3]

ABSTRACT: Data Analysis helps businesses to obtain pertinent and accurate information, which is useful for creating business planning, marketing strategies, and realigning the company's vision and mission. One of the prominent features to improve company growth is to analyze the user review. Review analysis is vital that gives a complete insight into what users are saying about the product purchased, website used, application, or overall experience of the service provided. It is fundamentally a customer-focused endeavour. It gives the ability to decide what to improve by taking customer values for consideration. The article discusses the examination of customer reviews provided by airlines. Online tools also assist organizations in understanding how different kinds of clients have varied objectives and how their choices impact reviews. In this study, we anticipate the analysis accuracy using several types of machine learning methods. This analysis presents airline companies to detect and fix service flow issues in addition to assisting customers in choosing an appropriate airline. The primary objective is to evaluate and compare the performance of diverse algorithms in analysis and satisfaction prediction.

KEYWORD: Sentiment Analysis, Natural Language Processing(NLP), Machine Learning Algorithms, Predictive Modelling.

I. INTRODUCTION

Feedback from customers is a potent source of information that can help you make more money or reduce marketing costs by providing vital insights into every aspect of the company. Customers and brands can develop trust by reading positive internet reviews. These evaluations can demonstrate that the business's products and services live up to expectations, satisfy client requirements, and accurately reflect its online representation. Review sites can also be used to promote or develop client loyalty initiatives. A new method for marketing and communication has been made possible by online feedback, which bridges the gap between conventional referrals and the increasingly prevalent form of opinion. Businesses will be able to enhance customer service by utilizing this knowledge as input.

In the airline business, providing excellent customer service has become crucial. Airlines put a lot of labour and effort into providing for the demands of their passengers and ensuring their safety. Customers that are dissatisfied tend to be disengaged, which results in fewer passengers and lower revenue. Customers should have a good experience from the moment they order their tickets until they board the aircraft, enter the airport, purchase meals there, check in, and arrive at their desired destinations. Airlines can discover areas for improvement and make changes to pricing and other policies by paying attention to customer input regarding the value for money of their products. Airlines can maintain their competitiveness and attract and retain devoted customers by analyzing customer feedback.

Machine learning can be used to customize the experience for customers. For example, it can be used to suggest trip destinations or flights based on past behavior or preferences. This can boost passenger satisfaction and loyalty while bringing in more money for the airline. Airlines use machine learning for a wide range of purposes, from enhancing passenger satisfaction to streamlining crew scheduling and minimizing maintenance problems. Machine learning is expected to become more and more important in the aviation business as data becomes more widely available. Popular Machine Learning algorithms used in this article are Support Vector Machine, Random Forest and Naïve Bayes.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. RELATED WORK

Jain et. al. [1] presents an investigated study airline recommendation prediction using machine learning algorithms. The text dataset was taken as an input and processed with Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbour (KNN). SVM results in high accuracy of 85.94%.

Akhmad et.al. [2] compares the effectiveness of Machine Learning algorithms such as Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Adaboost, Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), and Random Forest (RF) by dividing the Twitter airline sentiment data into positive, neutral, or negative categories. It also includes demonstration to prove that SMOTE oversampling can improve sentiment analysis accuracy with 97%.

Hasib K M [3] analyzed three machine learning algorithms (Decision Tree, Random Forest, XGBoost) and three deep learning algorithms (CNN, LSTM, BERT) for prediction. Further investigated how the model transform, feature extraction, and the number of classes influence classification outcomes. Finally the author concludes that BERT gives the best accuracy of 83% among all those classifiers.

EDA [4] investigated travelers' recommendation system using Convolutional Neural Network (CNN) model. Evaluation matrixes provided proves that the model presented exhibits better results. Further extended to prove how the quantity of features extracted affects performance.

III. DATASET DESCRIPTION

The dataset was taken from the public repository Kaggle [5], this data set describes the details of airline customer reviews. The data provided comes from passengers or travellers who have given their reviews regarding the airlines they have used. The dataset has 3700 records with 19 attributes that contains reviews of different facilities provided by the airline from 9th October 2011 to 19th November 2023. Table 1. shows the description of attributes and its data types in the dataset.

Figure 1. describes the type of the travel and the rating given by the customers for value for money. Figure 2. presents the data in graphical format that shows the overall rating of the customers with food and beverages. Figure 3. shows the rating of customers for cabin staff service respective to the date they flew. Figure 4. exhibits the seat type chosen by different types of customers.

S.No.	Attribute Name	Data Type	Description of Data
1	OverallRating	Numeric	Rating given by the customers in
2	ReviewHeader	String	Reviews given by the customers
3	Name	String	Name of the customers
4	Datetime	Date Time	Date of the travel
5	VerifiedReview	Boolean	True or False values of verified customers
6	ReviewBody	String	Descriptions of the customers review
7	TypeOfTraveller	String	Business, Family, Couple Leisure
8	SeatType	String	Economy Class, Economy Class

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711050

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

9	Route	String	Destination of the customers
10	DateFlown	Date Time	Date of travel
11	SeatComfort	Numeric	Customer ratings on seat comfort
12	CabinStaffService	Numeric	Customer ratings on Cabin Staff Service
13	GroundService	Numeric	Customer ratings on Ground Service
14	ValueForMoney	Numeric	Customer ratings on Value for money
15	Recommended	String	Yes or no review of customers
16	Aircraft	String	Name of the aircraft
17	Food&Beverages	Numeric	Customer ratings on Food & Beverages
18	InflightEntertainment	Numeric	Customer ratings on Inflight Entertainment
19	Wifi&Connectivity	Numeric	Customer ratings onWifi& Connectivity

Table 1. Dataset Description

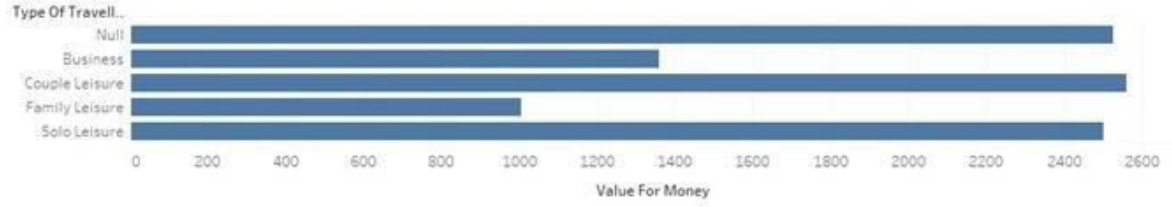

Fig1. Customer Rating

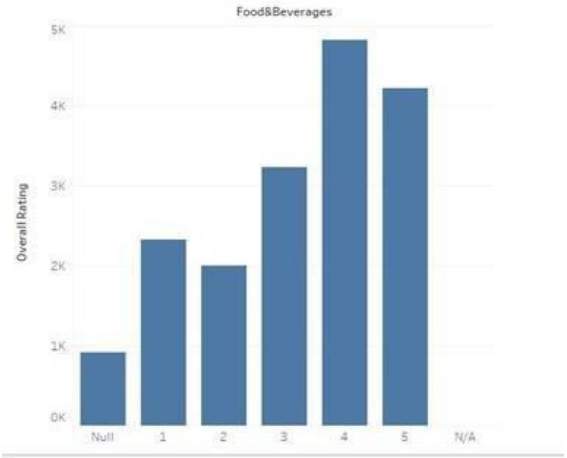

Fig 2. Food and Beverages rating

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711050

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

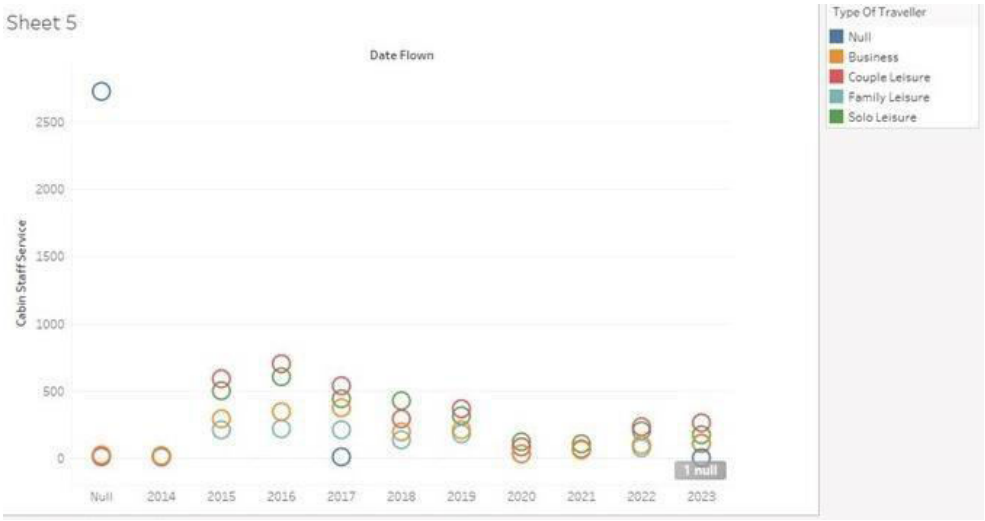

Fig. 3 Customer Rating for Cabin Staff Service

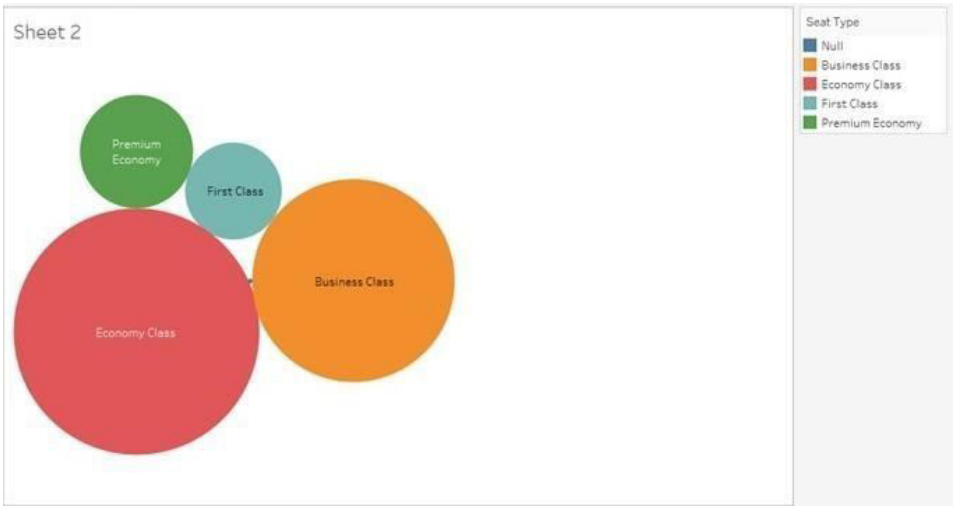

Fig 4. Seat Type – Customer Preferences

IV. METHODOLOGY

Naive Bayes:
A key method in machine learning for classification tasks, particularly in text classification and natural language processing, is the naive Bayes algorithm. With the assumption of feature (or predictor) independence within the dataset, Naive Bayes is based on the Bayes theorem [5]. The reason it is called "naive" is that it makes the assumption that a feature's existence in a class is independent of the existence of any other feature, an assumption that may not hold true in realworld situations yet frequently performs effectively in practice [8].

Random Forest:
A popular ensemble learning technique in machine learning for both classification and regression applications is the Random Forest algorithm [6]. In order to produce more reliable and accurate findings, it builds several decision trees during the training phase and then combines their predictions.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SVM:

Although it may also be used for regression and outlier identification, Support Vector Machines (SVM) is a supervised machine learning method that is mainly utilized for classification tasks [7].It works well with datasets that are both linearly and non-linearly separable.

V. RESULTS AND DISCUSSION

The dataset on airline passenger satisfaction was obtained from Kaggle. The dataset has already been divided into training and testing sets.The training dataset will be used for analysis and modeling, and the testing dataset will be used to evaluate the accuracy of our models. For accuracy prediction, we employed Random Forest, SVM, and Naive Bayes. When using Random Forest for prediction, the review dataset yields superior results compared to other methods. The better results are obtained because of a strong correlation between a label and the qualities. The following Table 2. presents the accuracy comparison among the three algorithms. The same results are depicted in the Figure 5. below.

Table 2. Accuracy Measure

Algorithm	Accuracy
Naïve Bayes	73.27%
SVM	75.36%
RF	81.57%

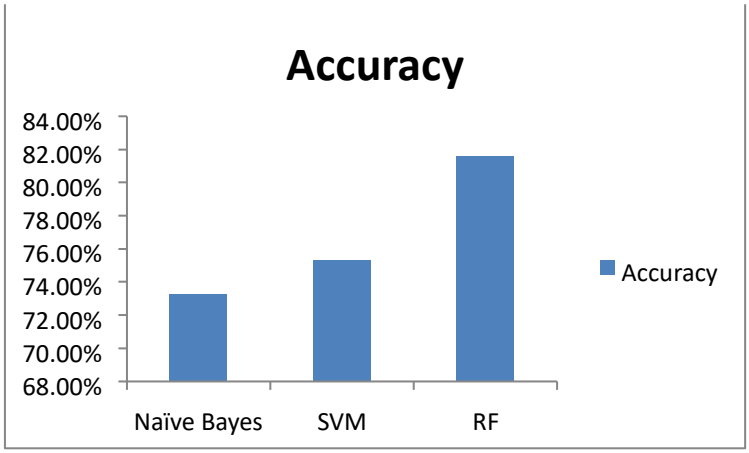

Fig 5. Accuracy Measure

VI. CONCLUSION

In this study, we looked into the potential uses of user-generated online reviews. Using user-generated web reviews, it also determines what characteristics influence preferences within various airline classes. Through this study, we can anticipate the satisfaction of the customer and conduct an exploratory analysis of reviews. In conclusion, we acknowledgethat our research examines a sample and offers insight into the review by offering explanatory analysis and forecasting the user's level of satisfaction. In future, we plan to assess the sentiment analysis of user reviews on the same dataset.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. Jain, P. K., Pamula, R., Ansari, S., Sharma, D., & Maddala, L. (2019, November). Airline recommendation prediction using customer generated feedback data. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 376-379). IEEE.
2. Akhmad, E. P. A., Adi, K., & Widodo, A. P. (2023). Machine learning approach to customer sentiment analysis in twitter airline reviews. In E3S Web of Conferences (Vol. 448, p. 02044). EDP Sciences.
3. Hasib, K. M. (2022). Sentiment analysis on Bangladesh airlines review data using machine learning (Doctoral dissertation, Brac University).
4. EDA, C. (2022). Unveiling the Ambivalence from the Airline Reviews: The Airline Recommendation System using CNN.
5. Huang, Y., & Li, L. (2011, September). Naive Bayes classification algorithm based on small sample set. In 2011 IEEE International conference on cloud computing and intelligence systems (pp. 34-39). IEEE.
6. Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3 (pp. 246-252). Springer Berlin Heidelberg.
7. Zhang, Y. (2012). Support vector machine classification algorithm and its application. In Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings, Part II 3 (pp. 179-186). Springer Berlin Heidelberg.
8. Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics, 403, 412.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com