**Title Page 1**

Automated Aero Assist Recommendation using Random Forest and Compared with Logistic Regression with Improved Accuracy

K. Varshini1, K.Sashi Rekha 2

K. Varshini1

Project student,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, Pincode: 602105.

kethuvarshini1556.sse@saveetha.com

K.Sashi rekha 2

Research Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, Pincode: 602105.

sashirekhak.sse@saveetha.com

**Keywords:** Novel Random Forest, Logistic Regression, Airline Reservation System, Airline, Travel, Flight, Ticket, Reservation.

**ABSTRACT**

**Aim:** The objective of this project is to enhance the accuracy of airline flight information systems. The goal is to improve customer-airline agency relationships by using the Airline Reservation System to streamline flight ticketing, selling, and air travel operations. **Materials and Methods:** This study evaluates Random Forest and Logistic Regression. 10,000 sample sets were analyzed using 10 iterations and an 80% G-power value. with a 95% confidence level, a programming language was used. A Kaggle dataset of 25,000 database Records from the field was used for the identification dataset. **Results:** The accuracy of the Novel Random forest is 96.57%. and accuracy of logistic Regression algorithm is 94.39%respectively, with a in the SPSS statistical analysis, a significant difference is defined as(p=0.016).(p<0.05) **Conclusion:** The results show that the Random forest is vastly superior to the logistic regression algorithm for airline reservation.

**Keywords:** Novel Random Forest, Logistic Regression, Airline Reservation System, Airline, Travel, Flight, Ticket, Reservation.

**INTRODUCTION**

The Airline Flight Information System was formerly independent and operated as separate systems. Each airline operated its own independent system, which was not integrated with that of other airlines or ticket agencies, and could only be accessed by a limited number of authorized airline personnel. In the 1970s, travel agents advocated for the ability to use the airlines' systems. The present-day process of managing air travel data involves the interconnection, storage, and retrieval of information via a network of Computer Reservations Systems. [(Wittmer, Bieger, and Müller 2011)](https://paperpile.com/c/MQ35Et/XHDX) The global distribution system facilitates the integration of ticket purchasing and sales across numerous airlines, while also enabling direct customer access to these systems. [(Kasabov and Warlow 2012)](https://paperpile.com/c/MQ35Et/64z5) Web-based global distribution system portals and gateways let users to directly buy tickets, choose seats, and make reservations for hotels and rental vehicles. [(Mckinsey Chief Marketing & Sales Officer Forum 2014)](https://paperpile.com/c/MQ35Et/0PKp) Crossing foreign borders empowers individuals to share evaluations on internet platforms, voice their thoughts, and significantly influences consumer interactions through reviews. Opinions may be stated in a single review (good or negative) or throughout reviews (conflicting). [(Smit et al. 2023)](https://paperpile.com/c/MQ35Et/soR5) Conflicting internet reviews, a lesser-studied issue, have gained popularity recently. These ratings help consumers choose an airline and help airlines uncover and fix service faults. [(Rana et al. 2019)](https://paperpile.com/c/MQ35Et/sdZX) We studied how airline evaluations' contradicting features cause passengers' feelings of ambivalence and indecision. [(National Research Council et al. 2011)](https://paperpile.com/c/MQ35Et/vKdP).The airline reservation applications include seat reservation and real-time information, enhancing customer experience and optimizing airline operations. It integrates with global distribution systems, supports multiple payment options, and provides analytics for informed decision-making in the airline industry.

This study explores autonomous driving's effects on policy and society, assesses relevant literature, and suggests further research. The ripple effect concept guides our review of automated vehicles' effects at three stages: first-order (traffic, travel cost, travel choices), second-order (vehicle ownership, location, land use, and transport infrastructure), and third-order. Google Scholar and Science Direct's databases of the most-cited research papers ended with 772 articles written about this area. Airline reservations systems originated in the late 1950s when American Airlines needed a real-time flight data system for all offices and automated booking and ticketing operations. Consequently, the Semi-Automated Business Research Environment was created and introduced in 1964. [(Cook and Billig 2017)](https://paperpile.com/c/MQ35Et/cpLD) Sabre's innovation was real-time inventory accuracy available to agents worldwide. Prior to this, manual methods need centralized reservation centers with physical cards representing inventory, such as aircraft seats. [(International Civil Aviation Organization 2004)](https://paperpile.com/c/MQ35Et/JfC2) The Airline Deregulation Act deregulated the airline business, requiring airlines to increase efficiency to compete in a free market. Previously, airlines received government-set rates to ensure profitability. The travel industry relied on the Airline Reservation System and its offspring in this unregulated climate. Early American commercial aviation had few passengers and strict regulations by the Civil Aeronautics Board on airline routes and pricing. [(Cusick, Cortes, and Rodrigues 2017)](https://paperpile.com/c/MQ35Et/QsyL). This excellent technique was used by airlines, Sheraton Hotels, and Goodyear for inventory management. The system was hindered by the necessity for local human operators to do lookups. [(Cheverton 2008)](https://paperpile.com/c/MQ35Et/Yp4J) Ticketing agents had to contact a booking office, where operators directed a small team to operate the Reservisor and read the findings over the phone. It was not possible for agents to directly query the system. [(International Association for Food Protection 2016)](https://paperpile.com/c/MQ35Et/hKMe) An Airline Reservation System is part of Passenger Service Systems, which enables direct passenger engagement. [(Krajangta, Mahidon, and Witthayāsāt 1997)](https://paperpile.com/c/MQ35Et/SzWq). A Computer Reservation System coordinates airline bookings using a Global Distribution System, enabling travel agents and other channels to book major airlines in one system. [(Ren 2000)](https://paperpile.com/c/MQ35Et/ayDV)

The research gap identified Online platforms are characterized differently. Online commerce, forums, blogs, social media, and review sites are examples of online communities that get reviews. Online ratings and reviews are often offered by customers having recent first-hand experience with a product or service (e.g., text-based opinion). Research indicates that online user reviews significantly impact product development. This study analyzes web reviews to assess airline critics' advice. This study helps service organizations determine how their services affect consumer marketing via online reviews. The research builds on previous results. Check whether internet reviews predict service recommendations. [(Joseph Pine, Pine, and Gilmore 1999)](https://paperpile.com/c/MQ35Et/2YSb) This paper proposes solutions to address ambivalence and provide advice to consumers and airlines. This study made two contributions: first, we used NLP approaches to preprocess traveler reviews in the recommender system. Second, the Convolutional Neural Network model was constructed to gather data from various social networks. [(Jayakumari et al. 2020)](https://paperpile.com/c/MQ35Et/yV6y).

**MATERIALS AND METHODS**

The training and testing of airline reservation systems are performed in the open source Laboratory,Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences. This group used in this study are two , a Random Forest, then a new Logistic Regression. Clinal determined sample size. The sample size was calculated using a 0.05 threshold, 80% G power, and 95% confidence range from previous studies.

Other features include a 512 MB Intel Core i5 CPU, Windows (version 11), Google Collab, and Jupyter Notebooks. IBM SPSS was used for the statistical evaluation in order to evaluate the precision. There are 25,000 rows and 24 columns of data used to train the algorithms, and the batch size of the data is 10,000. CSV files containing Identification Sex Customer Category Age, Travel Type, Class Flight, Distance,Wireless internet in flight, Good departure/arrival time, Booking online etc. this data were downloaded from the Kaggle website. SPSS is used to conduct a separate T-test study to compare the two methods for reliability.

**Random Forest**

Random Forest machine learning can do regression and classification. A decision tree "forest" is created by creating several decision trees. The Random Forest algorithm aggregates all tree projections to get the final forecast. The Random Forest approach is strong and frequently used because it can handle complicated non-linear relationships, missing data, and outliers and produce a probabilistic result estimate.

**Random forest Algorithm**

Step 1: Start by loading data.

Step 2: Divide the data into training and testing.

Step 3: Build a Random Forest classifier using hyperparameters like n\_estimators, maximum depth, etc.

Step 4: Train Random Forest classifier using data.

Step 5: Forecast using test results.

Step 6: Assess model accuracy, precision, and recall.

Step 7: Print or save model performance evaluation findings.

**Logistic Regression**

Logistic regression estimates a given example's class membership using supervised machine learning. Classification uses logistic regression. A sigmoid function estimates class membership probability from linear regression results, thus "regression". Logistic regression categorizes instances, whereas linear regression generates a constant result that may fluctuate.

**Logistic Regression Algorithm**

Step 1: Load your dataset (features X and labels Y).

Step 2: Divide the data into training and testing.

Step 3: Set hyperparameters like maximum depth and impurity measure (Gini index or entropy) for a Logistic Regression classifier.

Step 4: Train Logistic Regression on training data.

Step 5: Predict test data.

Step 6: Measure the model's performance (accuracy, precision, recall).

Step 7: Print or save model assessment findings to evaluate performance.

**Statistical Analysis**

Statistical analysis performed using the SPSS tool. The dependent variables in this study file include information such as Identification, Sex, Customer Category, Age, Travel Type, Class Flight, Distance, Wireless internet in flight, Good departure/arrival time, and online booking ticket with improved accuracy, which are influenced by the independent factors.

**RESULT**

Table 1 A novel Random Forest classifier that effectively complements a Logistic Regression technique. The table below displays the precision of Random Forest and Logistic Regression. This project aims to emphasize the importance and need of the Airline flight information system. is the process of dividing a dataset into smaller segments.

Table 2 presents a novel approach that combines the Random Forest and Logistic Regression algorithms to create a new classifier algorithm. T-tests comparing independent samples The table below presents the differences between the Novel Random Forest and Logistic Regression approaches in terms of significance, mean, and standard error. These differences were determined by SPSS analysis, with a p-value of 0.016, indicating statistical significance (p<0.05).

Table 3 presents the statistical data on the performance of the Logistic Regression and Random Forest algorithms. The results obtained from both methods were inputted into SPSS in order to calculate the mean, standard deviation, and standard error for both the Logistic Regression and Random Forest models. The accuracy of Logistic Regression is 96.57%, whereas the accuracy of Novel Random Forest is 94.39%.

The results indicate that the suggested approach has a success rate of 96.57% and Logistic Regression has a success rate of 94.39%. These figures are accompanied by a 95% confidence interval and a standard deviation of +/- 2 SD.

**DISCUSSION**

The Random Forest technique demonstrated greater performance compared to the Logistic Regression classifier in terms of accuracy (96.57%) and precision (94.39%). The study aim was to analyze the visitors' recommendation system by assessing reviews from diverse sources. The p-value is 0.016, which is less than 0.05. These trials indicate that Logistic Regression is less successful in sentiment detection. This paper suggests using a Random Forest as a potential solution.

The importance of the Airline management system lies in the automation of the organization's operations. It aids in the smooth functioning of the organization's dairy operations. The computerization of the organization's operations has had a favorable impact on its economics. [(Wittmer, Bieger, and Müller 2011)](https://paperpile.com/c/MQ35Et/XHDX) The study results will assist management in enhancing revenue creation and ensuring the efficient operation of daily operations. This presentation will be advantageous for anyone who use Airline flight information systems (Airline Reservation System), flight operators, air travel operators, travel agents, and airline agencies.[(Whiteley 2017)](https://paperpile.com/c/MQ35Et/1ivq) Customers use internet platforms to express their emotions. These emotions may be categorized as either good or negative, and at times, they can be a combination of both, resulting in mixed or contradicting evaluations. Customers have the ability to provide information, thoughts, and feedback about goods, services, and enterprises via online platforms such as Twitter, Instagram, and e-commerce sites. [(Davison et al. 2023)](https://paperpile.com/c/MQ35Et/pN7F) From a study standpoint, when airline enterprises and consumers engage in bivalent (both positive and negative), inconsistent, or contentious behaviors, there is an increased likelihood that customers may inaccurately interpret information. They elicit emotions of ambiguity and ambivalence in their demeanor. [(Christou 2006)](https://paperpile.com/c/MQ35Et/7ai0) Thus, we contend that the simultaneous presence of paradoxes in an object, such as online evaluations, generates attitudinal ambivalence in customers. This ambivalence arises from the item instilling in them an uncertain and conflicting attitude. Consequently, consumers are unable to make a choice and become perplexed. [(U.S. Department of Health and Human Services 2019)](https://paperpile.com/c/MQ35Et/HRmR)

Empirical research supports the relationship between e-service quality, perceived value, and e-loyalty in an electronic setting. Additionally, I examined the presence of a perceived value-satisfaction-loyalty link in relation to a tourist experience. The findings indicate that perceived value has a favorable and substantial influence on contentment, as does satisfaction on loyalty. However, it is acknowledged that more investigation is required. [(Gale 2009)](https://paperpile.com/c/MQ35Et/UTNf) Furthermore, it will benefit those in computer-related fields who want to have a comprehensive understanding of the system, as well as those doing research on related subjects. [(Marwedel 2021)](https://paperpile.com/c/MQ35Et/Drjr).

**CONCLUSION**

The research aimed to analyze tourist recommendation systems using Random Forest, employing metrics like F-measure, Precision, and Recall. The study assessed the system's functioning, feature extraction process, and the impact of extracted features on performance across N = 10 iterations. Random Forest achieved a classification accuracy of 96.57%, outperforming Logistic Regression (94.39%), with statistical significance at p=0.016 (p<0.05), indicating its superior examination result.

**DECLARATIONS**

**Conflicts of interest**

There is no reported conflict of interest in this study.

**Author's Contribution**

Author KV was responsible for collecting data, analyzing the data, and composing the manuscript of Natural Language Processing. Author KSR provided conceptual direction, supervised the work, and thoroughly reviewed the manuscript.

**Acknowledgements**

The authors extend their appreciation to the Saveetha School of Engineering and Saveetha Institute of Medical and Technical Science (formerly Saveetha University) for offering the necessary support to successfully complete this project.

**Funding**

The authors express their gratitude to the following organizations for their financial support, which made it possible to carry out the study.

1. CK technology Pvt. Ltd Chennai.
2. Saveetha School of Engineering.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha University.

**REFERENCES**

[Cheverton, Peter. 2008. *Global Account Management: A Complete Action Kit of Tools and Techniques for Managing Key Global Customers*. Kogan Page Publishers.](http://paperpile.com/b/MQ35Et/Yp4J)

[Christou, Anastasia. 2006. *Narratives of Place, Culture and Identity: Second-Generation Greek-Americans Return “Home.”* Amsterdam University Press.](http://paperpile.com/b/MQ35Et/7ai0)

[Cook, Gerald N., and Bruce G. Billig. 2017. *Airline Operations and Management: A Management Textbook*. Routledge.](http://paperpile.com/b/MQ35Et/cpLD)

[Cusick, Stephen K., Antonio I. Cortes, and Clarence C. Rodrigues. 2017. *Commercial Aviation Safety, Sixth Edition*. McGraw Hill Professional.](http://paperpile.com/b/MQ35Et/QsyL)

[Davison, Michelle, Jason Chan, Meg Clarke, Caroline Mitchell, Alan Yan, and Elwyn Henaway. 2023. “Yarning to Reduce Take Own Leave Events in First Nations Patients Presenting to the Emergency Department-Presenting the Qualitative Themes and Co-Design of the Deadly RED Project.” *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals*, December. https://doi.org/](http://paperpile.com/b/MQ35Et/pN7F)[10.1002/hpja.835](http://dx.doi.org/10.1002/hpja.835)[.](http://paperpile.com/b/MQ35Et/pN7F)

[Gale, Bradley. 2009. *Managing Customer Value: Creating Quality and Service That Customers Can See*. Simon and Schuster.](http://paperpile.com/b/MQ35Et/UTNf)

[International Association for Food Protection. 2016. *Procedures to Investigate Waterborne Illness*. Springer.](http://paperpile.com/b/MQ35Et/hKMe)

[International Civil Aviation Organization. 2004. *Manual on the Regulation of International Air Transport*.](http://paperpile.com/b/MQ35Et/JfC2)

[Jayakumari, J., George K. Karagiannidis, Maode Ma, and Syed Akhter Hossain. 2020. *Advances in Communication Systems and Networks: Select Proceedings of ComNet 2019*. Springer Nature.](http://paperpile.com/b/MQ35Et/yV6y)

[Joseph Pine, B., Joseph Pine, and James H. Gilmore. 1999. *The Experience Economy: Work Is Theatre & Every Business a Stage*. Harvard Business Press.](http://paperpile.com/b/MQ35Et/2YSb)

[Kasabov, E., and A. Warlow. 2012. *The Compliance Business and Its Customers: Gaining Competitive Advantage by Controlling Your Customers*. Springer.](http://paperpile.com/b/MQ35Et/64z5)

[Krajangta, Jurairat, Mahāwitthayālai Mahidon, and Mahāwitthayālai Mahidon Khana Witthayāsāt. 1997. *Self Service Airline Reservation System VIA Internet*.](http://paperpile.com/b/MQ35Et/SzWq)

[Marwedel, Peter. 2021. *Embedded System Design: Embedded Systems Foundations of Cyber-Physical Systems, and the Internet of Things*. Springer Nature.](http://paperpile.com/b/MQ35Et/Drjr)

[Mckinsey Chief Marketing & Sales Officer Forum. 2014. *Big Data, Analytics, and the Future of Marketing and Sales*. Createspace Independent Pub.](http://paperpile.com/b/MQ35Et/0PKp)

[National Research Council, Division of Behavioral and Social Sciences and Education, Board on Behavioral, Cognitive, and Sensory Sciences, and Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security. 2011. *Intelligence Analysis: Behavioral and Social Scientific Foundations*. National Academies Press.](http://paperpile.com/b/MQ35Et/vKdP)

[Rana, Nripendra P., Emma L. Slade, Ganesh P. Sahu, Hatice Kizgin, Nitish Singh, Bidit Dey, Anabel Gutierrez, and Yogesh K. Dwivedi. 2019. *Digital and Social Media Marketing: Emerging Applications and Theoretical Development*. Springer Nature.](http://paperpile.com/b/MQ35Et/sdZX)

[Ren, Dazhou. 2000. *An Online Airline Reservation System with XML*.](http://paperpile.com/b/MQ35Et/ayDV)

[Smit, Chloé C. H., Maarten Lambert, Kris Rogers, Steven P. Djordjevic, Antoine M. Van Oijen, Caitlin Keighley, Katja Taxis, Hamish Robertson, and Lisa G. Pont. 2023. “One Health Determinants of Antimicrobial Resistance in Humans in the Community: An Umbrella Review.” *International Journal of Molecular Sciences* 24 (24). https://doi.org/](http://paperpile.com/b/MQ35Et/soR5)[10.3390/ijms242417204](http://dx.doi.org/10.3390/ijms242417204)[.](http://paperpile.com/b/MQ35Et/soR5)

[U.S. Department of Health and Human Services. 2019. *TIP 35: Enhancing Motivation for Change in Substance Use Disorder Treatment (Updated 2019)*. Lulu.com.](http://paperpile.com/b/MQ35Et/HRmR)

[Whiteley, David. 2017. *An Introduction to Information Systems*. Bloomsbury Publishing.](http://paperpile.com/b/MQ35Et/1ivq)

[Wittmer, Andreas, Thomas Bieger, and Roland Müller. 2011. *Aviation Systems: Management of the Integrated Aviation Value Chain*. Springer Science & Business Media.](http://paperpile.com/b/MQ35Et/XHDX)

**TABLES AND FIGURES**

**Table 1.** Accuracy comparison between Logistic Regression and Random Forest Classifier. The following table displays the precision of Random Forest and Logistic Regression.

|  |  |  |
| --- | --- | --- |
| **, mki8S. No** | **Random Forest** | **Logistic Regression** |
| 1 | 96.16 | 94.72 |
| 2 | 96.61 | 94.93 |
| 3 | 96.02 | 94.95 |
| 4 | 96.99 | 94.04 |
| 5 | 96.34 | 94.79 |
| 6 | 96.76 | 94.03 |
| 7 | 96.93 | 94.71 |
| `8 | 96.97 | 94.18 |
| 9 | 96.83 | 94.93 |
| 10 | 96.66 | 94.35 |

|  |
| --- |
|  |

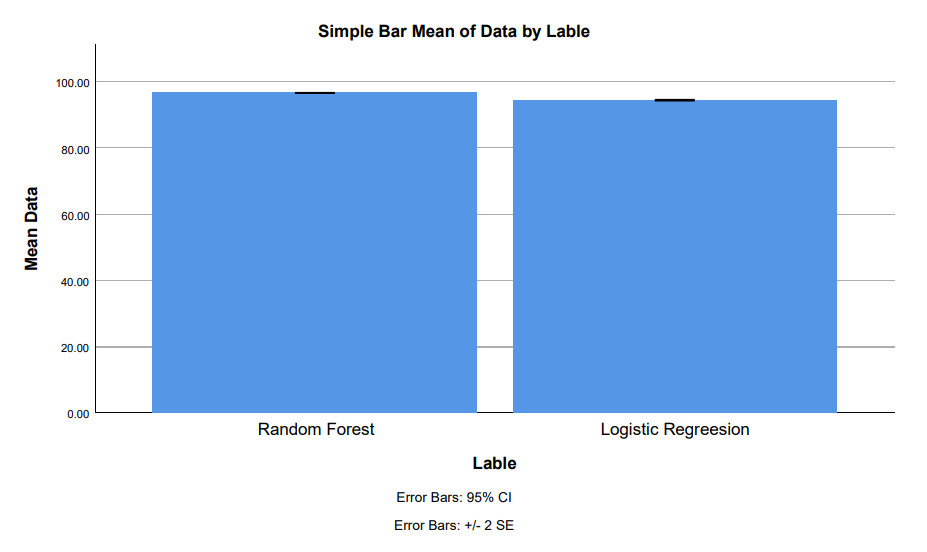
**Table 2**. Random Forest using Logistic Regression classifier and independent sample T-tests. The table below presents a comparison of the significance, mean, and standard error differences

between the Novel Random Forest and Logistic Regression approaches. These comparisons were made using SPSS analysis, with a significance level of p = 0.016 (p<0.05).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Levene’s Test for Equality of Variances** | | **T-Test for Equality of Mean** | | | | | **95%Confidence**  **Interval of Difference** | |
|  | **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean Difference** | **Std. Error Difference** | **Lower** | **Upper** |
| Equal variances assumed | 7.02 | 0.016 | 22.05 | 18 | 0.000 | 2.175 | 0.098 | 1.968 | 2.383 |
| Equal variances assumed |  |  | 22.05 | 13.64 | 0.000 | 2.175 | 0.098 | 1.963 | 2.387 |

**Table 3.** The method outputs for both Random Forest and Logistic Regression were used in SPSS to get the mean, standard deviation, and standard error. The Random Forest model achieves an accuracy of 96.57%, whereas the Logistic Regression model achieves an accuracy of 94.39%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Algorithm** | **N** | **Mean** | **Std.Deviation** | **Std.Error Mean** |
| **Accuracy** | Novel Random Forest | 10 | 96.57 | 0.145 | 0.046 |
| Logistic Regression | 10 | 94.39 | 0.275 | 0.087 |

****

**Fig. 1.** Statistical analysis was conducted using the SPSS tool to determine the accuracy of the data. The findings of the suggested method yielded an accuracy of 96.57% using Random Forest. This accuracy was compared to a baseline accuracy of 94.39% which is Logistic Regression, with an error bar of 95% confidence interval and a standard deviation of +/- 2.