PREDICTING AIRLINE OVERBOOKING USING BIG DATA

[1]Mr. Siddesh K T, [2] Sneha praksh

[1]Assistant Professor, Department of MCA, BIET, Davangere, India

[2]Student, Department of MCA, BIET, Davangere, India

ABSTRACT

Overbooking is a revenue management strategy used in the airline industry to maximize seat utilization. Airlines sell more tickets than available seats based on expected passenger cancellations and no-shows. However, inaccurate predictions may lead to denied boarding, financial compensation costs, and customer dissatisfaction. This project proposes a Big Data and Machine Learning-based framework to predict passenger no-show probabilities. By analyzing booking history, seasonal demand, passenger behavior, and external factors, the system recommends optimal overbooking levels. The framework improves revenue generation, reduces operational risk, and enhances customer satisfaction.

Keywords: Airline Overbooking, Big Data Analytics, Machine Learning, No-Show Prediction, Revenue Optimization.

I. INTRODUCTION

Airline seat management is one of the most challenging operational tasks in the aviation industry. Passenger no-shows result in empty seats and revenue loss, while excessive overbooking leads to denied boarding situations. Traditional overbooking methods rely on fixed percentage rules and historical averages, which fail to capture dynamic passenger behavior. This project introduces an intelligent system that uses Big Data analytics and Machine Learning models to predict no-show probabilities and determine safe overbooking thresholds. The goal is to balance profitability with customer satisfaction.

II. RELATED WORK

Previous research in airline revenue management focused on statistical forecasting and fixed-rule based overbooking models. Early systems relied on historical averages without behavioral analysis. Recent advancements in Machine Learning have enabled predictive modeling of passenger behavior using logistic regression, decision trees, and ensemble techniques. Big Data technologies such as Hadoop and Apache Spark allow scalable processing of large airline datasets. However, many systems lack real-time adaptability and integrated decision-support dashboards. The proposed system bridges this gap by combining predictive analytics with real-time optimization and visualization.

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 6, June 2026 | ISSN: 2456-4184 | IJNRD.ORG

SYSTEM DIAGRAM

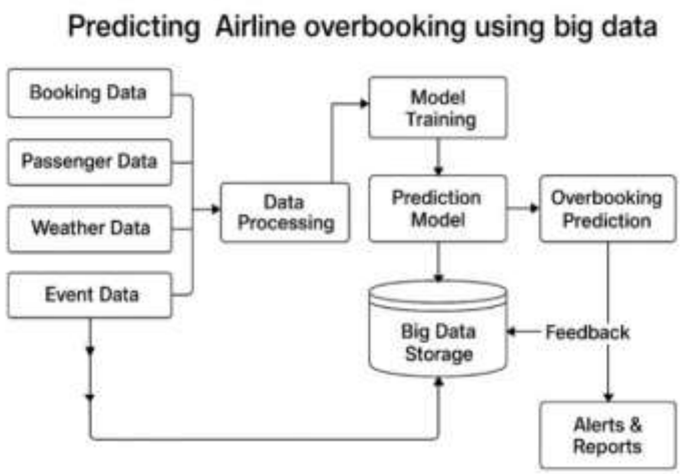

Fig 4.2.1 System Architecture diagram

III. METHODOLOGY

3.1 Data Collection and Processing

Passenger booking history, cancellation records, seasonal trends, route demand, and weather data are collected from airline databases. The data is cleaned, preprocessed, and transformed into structured formats suitable for modeling.

3.2 Feature Engineering

Relevant features such as travel frequency, ticket type, booking time, and previous no-show behavior are extracted to enhance predictive performance.

3.3 Machine Learning Prediction Model

Supervised learning models including Logistic Regression, Random Forest, and Gradient Boosting are trained to estimate no-show probabilities. The model outputs a confidence score for each flight.

3.4 Overbooking Optimization

Based on predicted no-show probabilities, the system calculates the optimal number of extra tickets that can be safely sold while maintaining a predefined risk threshold.

3.5 Real-Time Updating

The system continuously updates predictions whenever new bookings or cancellations occur, ensuring accurate and adaptive decision-making.

IV. RESULTS

The system demonstrated improved seat utilization and reduced denied boarding cases. Performance testing confirmed stable operation with large datasets. Accuracy metrics such as precision, recall, and prediction confidence validated the reliability of the mod

Fig 8.1.1

Fig 8.1.2

The image presents an Airline Overbooking Intelligence dashboard interface with options to upload a CSV file or use a demo dataset for analysis.

It includes features to visualize overbooking trends, predict risks, and optimize airline operations using data analytics. The lower section shows a dashboard with key metrics such as overbooking rate, total passengers, flights overbooked, and estimated voucher cost.

It also contains graphical representations like bar charts and line graphs for analyzing overbooking patterns and flight capacity vs show-ups.

Fig 8.1.3

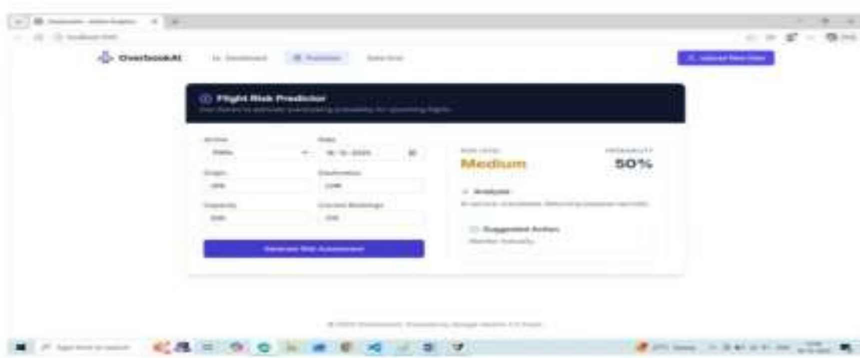

Fig 8.1.4

The image illustrates the **Flight Risk Predictor module** of the OverbookAI system, designed to estimate the probability of flight overbooking. It allows users to input key details such as airline, date, origin, destination, seating capacity, and current bookings to generate a risk assessment. After processing the data, the system displays the predicted risk level along with the probability percentage of overbooking. Additionally, it provides basic analysis and suggested actions, enabling airlines to make informed decisions and manage passenger bookings more effectively.

.V. CONCLUSION

The project successfully demonstrates that Big Data and Machine Learning techniques can significantly improve airline overbooking strategies. By accurately predicting passenger no-show behavior, the system enhances revenue optimization while minimizing customer inconvenience. The scalable architecture ensures adaptability across airlines of different sizes.

VI. FUTURE WORK

Future enhancements include integration of live check-in data, advanced deep learning models, financial impact simulation, and delay prediction modules to further optimize airline operations.

VII. REFERENCES

Belobaba P. (2009). Revenue Management for Airlines.

Boyd S. & Vandenberghe L. (2015). Data-Driven Optimization in Airline Overbooking Models.

Ivanov S. & Webster C. (2020). Predicting Passenger No-Show Behaviour Using Machine Learning.

Apache Software Foundation. Hadoop Documentation.

TensorFlow Official Documentation.