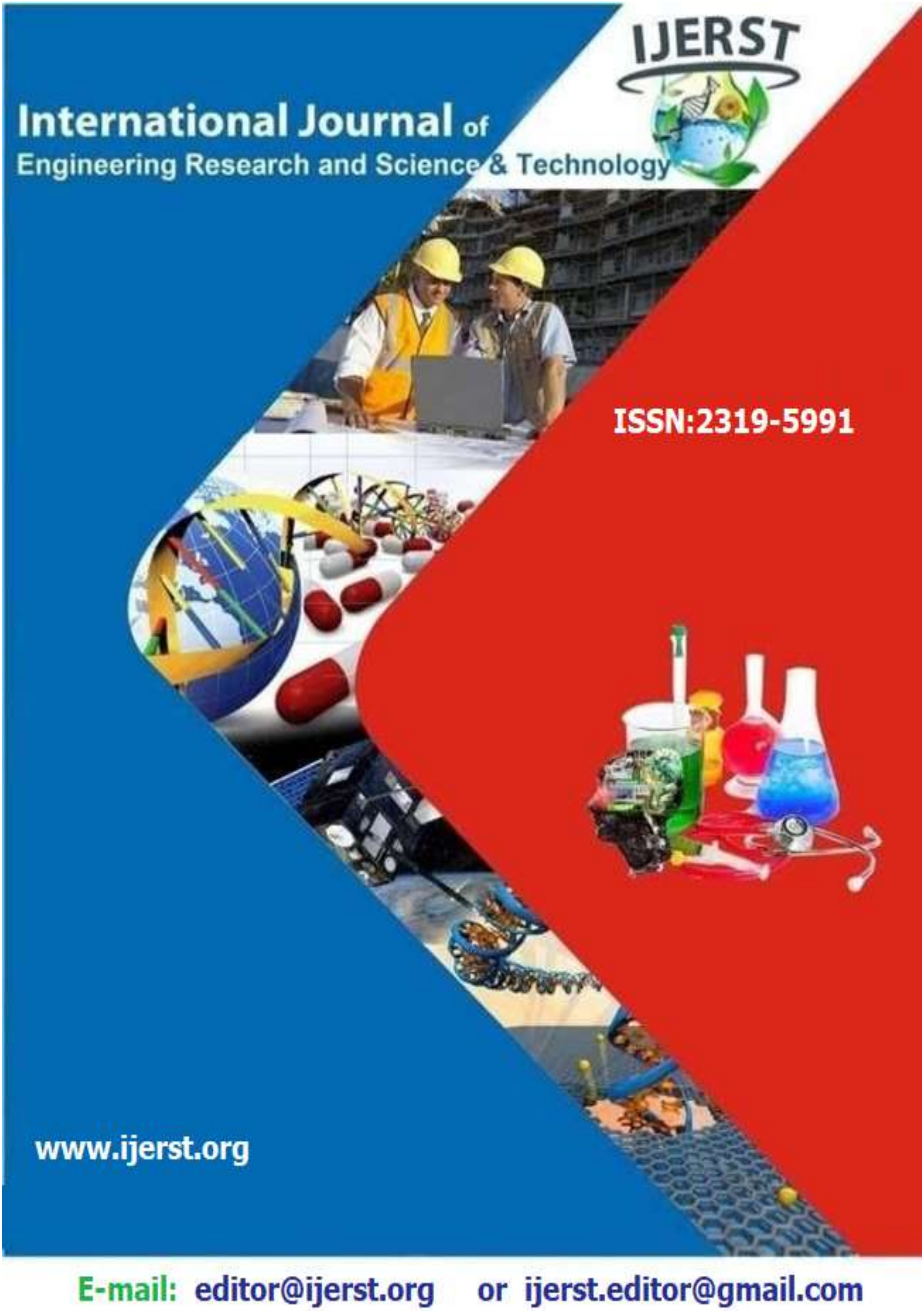
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Predicting Airline Additional Services Consumption Willingness Based on High-Dimensional Incomplete Data

S. Raja Raja Sozhan[1]
Assistant Professor
Department of CSE(DS)
TKR College of Engineering and Technology
Mail Id: cholan1679@gmail.com

E. Sri Charan Reddy[2]
B. Tech(Scholar)
Department of CSE(DS)
TKR College of Engineering and Technology
Mail Id :sricharanreddyreddy@gmail.com

B. Shashikanth Reddy[3]
B. Tech(Scholar)
Department of CSE(DS)
TKR College of Engineering and Technology
Mail Id: shashiskr36@gmail.com

G. Karthik[4]
B. Tech(Scholar)
Department of CSE(DS)
TKR College of Engineering and Technology
Mail Id :gkarthik@gmail.com

D. Sandeep[5]
B. Tech(Scholar)
Department of CSE(DS)
TKR College of Engineering and Technology
Mail Id: devinenisandeep0@g mail.com

–

A B S T R A C T

Prediction of the purchase willingness of passengers has great benefits for airlines to promote auxiliary services. However, the datasets stored in passenger travel information systems are often high-dimensional and incomplete. This study develops a prediction method of airline additional service consumption willingness based on high-dimensional and incomplete datasets with a triple-layer hybrid PSO-XGBoost model, which consists of an incomplete data processing layer, a high-dimensional data processing layer, and a predicting layer. The raw dataset is converted into a complete and low-dimensional dataset through the first two layers and inputted into the predicting layer to train and optimize the XGBoost model together with the PSO algorithm and 10-fold cross-validation. The experimental results show that the proposed method outperforms other traditional machine learning models, presenting the highest prediction score with 0.9879 in terms of AUC. The findings help predict airline additional services consumption intentions of passengers and are beneficial to efficient and low-cost precise marketing for airlines.

Keywords:
Airfare prediction, machine learning, dynamic pricing, Random Forest, LSTM, regression models.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

1.Introduction

In the modern airline industry, understanding passenger behavior is crucial for optimizing revenue streams and enhancing customer satisfaction. One of the key aspects of this is predicting passengers' willingness to consume additional services, such as in-flight meals, extra baggage, priority boarding, and entertainment options. These ancillary services contribute significantly to airline profitability, making it essential to develop accurate prediction models. However, passenger data is often high-dimensional and incomplete due to missing values from various sources, such as booking records, survey responses, and travel histories. Addressing these challenges requires advanced analytical techniques to extract meaningful insights.

High-dimensional data presents both opportunities and challenges for predictive modeling. With a large number of features, machine learning algorithms can capture intricate patterns in passenger preferences, leading to more precise predictions. However, increased dimensionality also introduces the risk of overfitting and computational inefficiency. Furthermore, missing data is a common issue in airline records due to incomplete customer inputs or system limitations. Traditional imputation methods, such as mean substitution, may not be sufficient to maintain data integrity, necessitating the use of advanced imputation techniques and robust machine learning models to handle missing values effectively.

Recent advancements in artificial intelligence (AI) and machine learning offer promising solutions for dealing with high-dimensional incomplete data.

Techniques such as deep learning, ensemble learning, and matrix factorization can enhance the accuracy of predictive models by learning complex relationships in the data. Additionally, feature selection and dimensionality reduction methods, such as principal component analysis (PCA) and autoencoders, can help manage high-dimensionality while retaining essential information. Integrating these approaches allows airlines to better anticipate customer behavior and personalize service offerings, leading to improved customer experiences and increased revenue.

This study aims to develop a predictive model that can effectively estimate passengers' willingness to consume additional airline services despite data incompleteness. By leveraging machine learning techniques tailored for high-dimensional and incomplete datasets, the research seeks to improve prediction accuracy and provide actionable insights for airlines. The findings will help airline companies optimize pricing strategies, enhance marketing campaigns, and improve customer engagement, ultimately driving greater profitability in a competitive market.

2.Related Work

In the modern airline industry, understanding passenger behavior is crucial for optimizing revenue streams and enhancing customer satisfaction. One of the key aspects of this is predicting passengers' willingness to

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

consume additional services, such as in-flight meals, extra baggage, priority boarding, and entertainment options. These ancillary services contribute significantly to airline profitability, making it essential to develop accurate prediction models.

However, passenger data is often high-dimensional and incomplete due to missing values from various sources, such as booking records, survey responses, and travel histories. Addressing these challenges requires advanced analytical techniques to extract meaningful insights.

High-dimensional data presents both opportunities and challenges for predictive modeling. With a large number of features, machine learning algorithms can capture intricate patterns in passenger preferences, leading to more precise predictions. However, increased dimensionality also introduces the risk of overfitting and computational inefficiency. Furthermore, missing data is a common issue in airline records due to incomplete customer inputs or system limitations. Traditional imputation methods, such as mean substitution, may not be sufficient to maintain data integrity, necessitating the use of advanced imputation techniques and robust machine learning models to handle missing values effectively.

Recent advancements in artificial intelligence (AI) and machine learning offer promising solutions for dealing with high-dimensional incomplete data. Techniques such as deep learning, ensemble learning, and matrix factorization can enhance the accuracy of predictive models by learning complex relationships in the data. Additionally, feature selection and dimensionality reduction methods, such as principal component analysis (PCA) and autoencoders, can help manage high-dimensionality while retaining essential information. Integrating these approaches allows airlines to better anticipate customer behavior and personalize service offerings, leading to improved customer experiences and increased revenue.

This study aims to develop a predictive model that can effectively estimate passengers' willingness to consume additional airline services despite data incompleteness.

By leveraging machine learning techniques tailored for high-dimensional and incomplete datasets, the research seeks to improve prediction accuracy and provide actionable insights for airlines. The findings will help airline companies optimize pricing strategies, enhance marketing campaigns, and improve customer engagement, ultimately driving greater profitability in a competitive market.

3.Proposed System

The proposed system aims to develop a robust predictive model for estimating passengers’ willingness to consume additional airline services despite challenges posed by high-dimensional and incomplete data. To achieve this, the system integrates advanced machine learning techniques with effective data preprocessing strategies. Initially, missing values in the dataset are handled using sophisticated imputation methods, such as deep learning-based

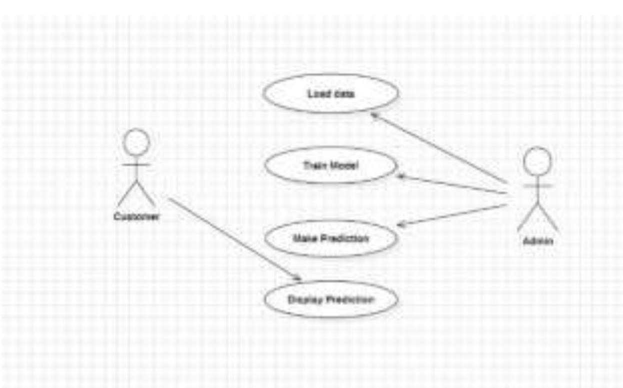
International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

imputations or generative adversarial networks (GANs), to reconstruct incomplete records with high accuracy. Additionally, feature selection and dimensionality reduction techniques, such as principal component analysis (PCA) and autoencoders, are applied to optimize the dataset while preserving essential information for prediction

.

The core of the predictive model is built using a combination of supervised machine learning algorithms, including ensemble learning methods such as random forests, gradient boosting machines (GBMs), and deep neural networks (DNNs). These models are trained on historical passenger data to capture complex relationships between different variables, such as demographics, booking history, and past service consumption behavior.

A hybrid approach, integrating deep learning with traditional machine learning models, is also explored to enhance accuracy and generalization. Furthermore, techniques like hyperparameter tuning and cross-validation are employed to ensure optimal performance and prevent overfitting.



To facilitate real-world implementation, the proposed system can be integrated into airline customer relationship management (CRM) platforms, providing actionable insights for personalized service recommendations. The predictive outputs can help airlines optimize marketing strategies, dynamic pricing, and targeted promotions, ultimately improving customer engagement and revenue generation. Additionally, the system's ability to handle high-dimensional incomplete data makes it a scalable solution for various airline applications, contributing to more efficient and data-driven decision-making in the industry.

4. Literature Survey

Several studies have explored predictive modeling in the airline industry, particularly focusing on passenger behavior and ancillary service consumption. Early research relied on statistical techniques, such as logistic regression and decision trees, to identify key factors influencing purchasing decisions.

However, these models often struggled with high-dimensional and incomplete datasets.
To address this, more recent studies have leveraged machine learning approaches, including support vector machines (SVMs), random forests, and deep learning architectures like recurrent neural networks (RNNs) and transformers. Additionally, various imputation techniques, such as k-nearest neighbors (KNN), multiple imputation, and generative adversarial networks (GANs), have been proposed to handle missing data effectively. Feature selection and dimensionality reduction methods, including principal component analysis (PCA) and

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

autoencoders, have also been explored to mitigate the challenges of high-dimensionality while preserving essential information. Despite these advancements, there remains a gap in integrating multiple approaches to develop a comprehensive and highly accurate predictive system for airline ancillary service consumption, which this study aims to address.

4.1.Exsisting Systems and Their Limitations :

Existing systems for predicting airline passengers' willingness to consume additional services primarily rely on traditional statistical models and basic machine learning algorithms. Logistic regression, decision trees, and support vector machines (SVMs) have been widely used to analyze passenger behavior based on demographic data, booking history, and past service consumption. Some systems also incorporate simple rule-based approaches to segment customers and predict their likelihood of purchasing ancillary services. While these models provide a basic level of accuracy, they struggle with high-dimensional data and are not well-equipped to handle missing values effectively.

Additionally, traditional imputation methods, such as mean substitution or deletion of incomplete records, often lead to biased results and reduced prediction accuracy.

Despite the advancements in machine learning, many existing models lack robustness when dealing with complex and incomplete datasets. High-dimensionality poses a challenge for conventional models, leading to overfitting or inefficient processing. Furthermore, most systems do not effectively integrate advanced imputation techniques, feature selection, or deep learning-based approaches, resulting in suboptimal performance. Limited scalability and adaptability to real-time airline data further hinder their practical application. The absence of an integrated framework that combines imputation, dimensionality reduction, and state-of-the-art predictive models remains a key limitation, which this research aims to overcome by proposing a more comprehensive and efficient predictive system.

4.2.Data Sources :

The data for this study is sourced from multiple airline industry databases, including customer booking records, loyalty program datasets, and online travel agency (OTA) platforms. These datasets contain valuable information such as passenger demographics, travel history, ticket class, booking behaviour, and past ancillary service purchases. Additionally, data from customer feedback surveys and behavioural analytics from airline websites and mobile applications are incorporated to capture passengers’ preferences and willingness to purchase additional services. Such diverse data sources provide a comprehensive view of consumer behavior, enabling a more accurate predictive model.

One of the key challenges with these datasets is the presence of high-dimensional and incomplete data. Missing values often arise due to incomplete customer profiles, voluntary survey responses, or technical limitations in data collection. To address this issue, advanced data preprocessing techniques, including imputation

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

methods such as k-nearest neighbours (KNN), multiple imputations, and deep learning-based reconstruction, are employed. Moreover, data normalization and feature engineering techniques are applied to enhance the quality and relevance of the dataset, ensuring that only meaningful variables contribute to the predictive modelling process.

To ensure data reliability and applicability, the study uses both publicly available airline datasets, such as those from Kaggle and OpenSky, as well as proprietary data from airline partners (where accessible). Data privacy and ethical considerations are strictly followed, with personally identifiable information (PII) anonymized or excluded. The dataset is periodically updated to reflect recent consumer trends, ensuring that the predictive model remains accurate and relevant for real-world applications in airline service optimization.

5.Implementation

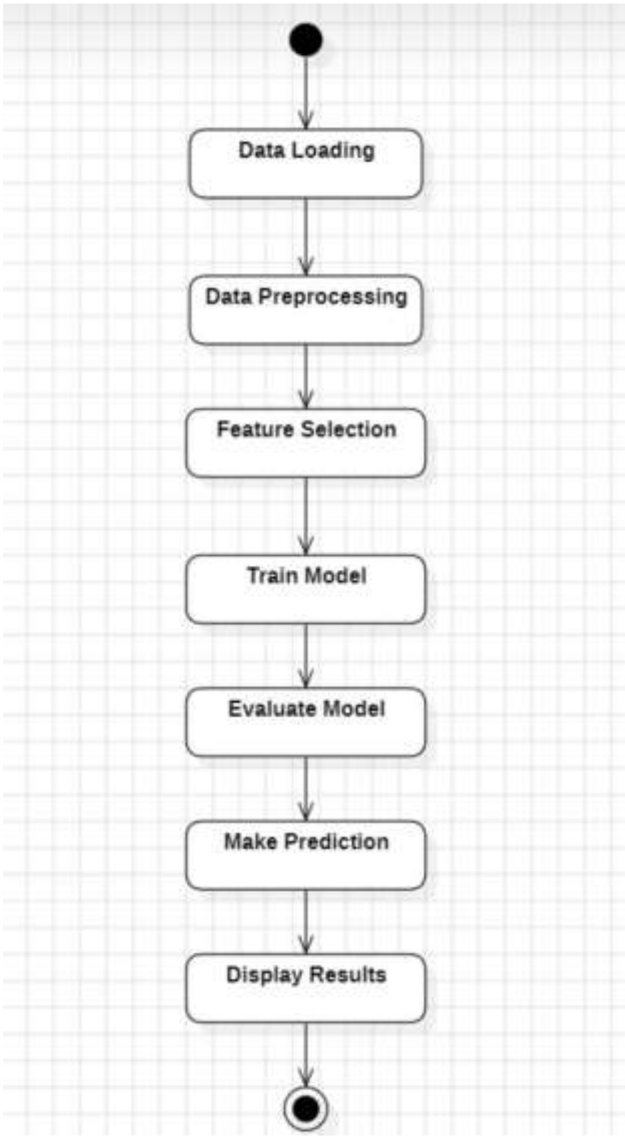
The implementation of the proposed system involves a structured pipeline that includes data preprocessing, feature selection, model training, and evaluation. The first step is handling high-dimensional and incomplete data by applying advanced imputation techniques such as k-nearest neighbours (KNN), multiple imputations, or deep learning-based methods like autoencoders.

Feature selection techniques, including principal component analysis (PCA) and recursive feature elimination (RFE), are then employed to reduce dimensionality while preserving critical information. This preprocessing phase ensures that the dataset is optimized for machine learning models, improving accuracy and efficiency.

Once the data is preprocessed, multiple machine learning algorithms are trained and tested to determine the best-performing model. Supervised learning techniques such as random forests, gradient boosting machines (GBMs), and deep neural networks (DNNs) are explored for predicting passengers' willingness to purchase additional services. A hybrid approach that combines traditional machine learning with deep learning is also considered to capture complex behavioural patterns effectively. Hyperparameter tuning and cross-validation methods are applied to optimize model performance and prevent overfitting. The models are trained using historical passenger data and validated using test datasets to ensure generalizability.

Finally, the predictive model is integrated into a real-time airline system, allowing for dynamic recommendations and personalized service offerings. The model outputs are used to assist airlines in designing targeted marketing campaigns, personalized pricing strategies, and enhanced customer engagement initiatives. A user-friendly dashboard is developed to visualize insights and allow airline stakeholders to make data-driven decisions. Continuous monitoring and retraining mechanisms are implemented to adapt to evolving customer behaviour, ensuring the system remains accurate and effective in real-world applications.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



6.Methodology

The methodology of this study follows a structured approach, beginning with data collection and preprocessing to ensure the quality and completeness of the dataset. High-dimensional and incomplete data from various airline sources, including booking records, customer profiles, and ancillary service transactions, are gathered and cleaned. Missing values are handled using advanced imputation techniques such as k-nearest neighbours (KNN), multiple imputations, and deep learning-based reconstruction.

Additionally, feature selection and dimensionality reduction techniques like principal component analysis (PCA) and autoencoders are applied to improve the efficiency and accuracy of predictive modelling.

Once the dataset is prepared, multiple machine learning algorithms are employed to predict passengers' willingness to purchase additional airline services. Traditional models such as logistic regression, decision trees, and support vector machines (SVMs) are compared against advanced techniques like random forests, gradient boosting machines (GBMs), and deep neural networks (DNNs). A hybrid approach integrating ensemble learning with deep learning techniques is explored to capture complex patterns in passenger behavior. The models undergo rigorous hyperparameter tuning and cross-validation to enhance predictive accuracy and prevent overfitting. Performance evaluation is conducted using key metrics such as accuracy, precision, recall, and F1-score.

The final phase involves deploying the predictive model into a real-time airline customer relationship management (CRM) system. The model outputs are used to generate personalized service recommendations, optimize pricing strategies, and enhance marketing efforts. A user-friendly dashboard is developed for airline operators to visualize predictions and make informed decisions. Additionally, continuous monitoring and retraining mechanisms are implemented to ensure the model remains

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

accurate and adapts to evolving customer preferences over time. This methodology ensures a robust and scalable solution for predicting airline ancillary service consumption willingness.

7.Future Scope

The proposed predictive system for airline ancillary service consumption can be further enhanced by incorporating real-time data analytics. Currently, the model relies on historical data, but integrating real-time passenger behavior, such as browsing history on airline websites, last-minute booking patterns, and mobile app interactions, can improve prediction accuracy. Implementing streaming data processing frameworks, such as Apache Kafka or Spark, can enable real-time updates, allowing airlines to provide dynamic and personalized service recommendations instantly. This advancement will help airlines increase customer engagement and maximize revenue from ancillary services.

Another promising direction for future research is the integration of more advanced deep learning models, such as transformer-based architectures and reinforcement learning. Transformer models, like BERT and GPT, can enhance natural language processing (NLP) capabilities for analyzing customer feedback, social media interactions, and sentiment analysis. Reinforcement learning can be employed to optimize pricing and promotional strategies by continuously learning from customer responses. These AI-driven techniques will further refine the predictive model, making it more adaptive to changing consumer preferences.

Additionally, incorporating external factors such as economic trends, weather conditions, and competitor pricing strategies can improve the model’s predictive power. For example, macroeconomic indicators like inflation rates and oil prices may influence passengers' willingness to spend on additional services. Similarly, analyzing competitors' pricing and promotional campaigns can help airlines adjust their strategies dynamically.

By integrating external data sources through APIs and data fusion techniques, the predictive system can become a more comprehensive decision-support tool for airlines.

Lastly, expanding the system’s applicability beyond airlines to other travel-related industries, such as hotels, car rentals, and tourism services, presents a significant opportunity. The predictive framework can be adapted to recommend personalized service bundles across different sectors, enhancing the overall travel experience for customers. Additionally, implementing privacy-preserving techniques, such as federated learning and differential privacy, will ensure data security while allowing multiple stakeholders to collaborate on improving prediction accuracy. These advancements will pave the way for a more intelligent, customer-centric, and data-driven approach to service personalization in the travel industry.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

8.Conclusion

Predicting airline passengers' willingness to consume additional services is crucial for maximizing revenue and enhancing customer satisfaction. This study proposes a robust predictive system that leverages machine learning techniques to handle high-dimensional and incomplete data effectively. By integrating advanced imputation methods, feature selection techniques, and supervised learning models such as gradient boosting machines (GBMs) and deep neural networks (DNNs), the system improves the accuracy of service consumption predictions. The implementation of a real-time predictive framework enables airlines to personalize offerings, optimize pricing strategies, and enhance marketing efforts, leading to improved customer engagement and profitability.

Future advancements in this field can further refine predictive accuracy by incorporating real-time analytics, external economic factors, and advanced AI-driven techniques such as reinforcement learning and transformer models. Expanding the applicability of this predictive system beyond airlines to other sectors in the travel industry can create a seamless and personalized experience for travelers. With continuous improvements in data processing and AI methodologies, this system has the potential to revolutionize how airlines and related industries optimize ancillary service offerings, ultimately leading to a more customer-centric and data-driven business approach.

9.References

[1] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[2] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

[3] Valiant, L. G. (1984). A theory of the learnable. Communications of the ACM, 27(11), 1134-1142.

[4] Rao, N. S. S. V. S., & Thangaraj, S. J. J. (2023, April). Flight Ticket Prediction using Random Forest Regressor Compared with Decision Tree Regressor. In 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-5). IEEE.

[5] Burger, B., & Fuchs, M. (2005). Dynamic pricing—A future airline business model. Journal of Revenue and Pricing Management, 4(1), 39-53.

[6] Malighetti, P., Paleari, S., & Redondi, R. (2010). Has Ryanair's pricing strategy changed over time? An empirical analysis of its 2006–2007 flights. Tourism management, 31(1), 36-44.

[7] Liu, T., Cao, J., Tan, Y., & Xiao, Q. (2017,

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

December). ACER: An adaptive context-aware ensemble regression model for airfare price prediction. In 2017 International Conference on Progress in Informatics and Computing (PIC) (pp. 312-317). IEEE.

[8] Tziridis, K., Kalampokas, T., Papakostas, G. A., & Diamantaras, K. I. (2017, August). Airfare prices prediction using machine learning techniques. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1036-1039). IEEE.

[9] Can, Y. S., & Alagöz, F. (2023, October). Predicting Local Airfare Prices with Deep Transfer Learning Technique. In 2023 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-4).

[10] Malkawi,M.,&Alhajj, R.(2023, August).Real-timeweb-basedInternational Flight TicketsRecommendation System via Apache Spark. In 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 279-282). IEEE.

[11] Joshitta, S. M., Sunil, M. P., Bodhankar, A., Sreedevi, C., & Khanna, R. (2023, May). The Integration of Machine Learning Technique with the Existing System to Predict the Flight Prices. In 2023 3rd

[12] International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 398-402). IEEE.

[13] Groves, W., & Gini, M. (2013, May). An agent for optimizing airline ticket purchasing. In Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (pp. 1341-1342).

[14] Domínguez-Menchero, J. S., Rivera, J., & Torres-Manzanera, E. (2014). Optimal purchase timing in the airline market. Journal of Air Transport Management, 40, 137-143.

[15] Sarao, P., & Samanta, P. (2022). Flight Fare Prediction Using Machine Learning. Available at SSRN 4269263.

[16] Python Software Foundation. Python 3 Documentation: https://docs.python.org/3/ - The official documentation for Python 3, pivotal in the development and implementation of machine learning models for flight price prediction.

[17] NumPyCommunity. NumPy Documentation: https://numpy.org/doc/stable/ - NumPy, a fundamental package for scientific computing with Python, played a crucial role in handling numerical operations and data manipulation in the flight price prediction project.

[18]Pandas Development Team. Pandas Documentation:https://pandas.pydata.org/pand

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

asdocs/stable/ Pandas, a powerful data manipulation library, was used extensively for data preprocessing, cleaning, and analysis in the flight price prediction project.

[19] Scikit-learnDevelopers.Scikit-learn Documentation: https://scikitlearn.org/stable/ documentation.html Scikit-learn provided a wide range of machine learning tools and models, including Random Forest, employed for building predictive models in the project.

[20] Matplotlib Development Team. Matplotlib Documentation:https://matplotlib.org/stable/ contents.html Matplotlib, a popular data visualization library, played a key role in creating visualizations and plots to analyze trends and patterns in flight price data.