# FORECASTING OF AIRLINE PASSENGERS BASED ON MACHINE LEARNING

Guide :- Dr. M.K.Demde
Guide, Department of Electronics and Telecommunications, Priyadarshini College of Engineering, Nagpur, India

Co-Guide :- Dr. A.S.Khobragade
Co-Guide, Department of Electronics and Telecommunications, Priyadarshini College of Engineering, Nagpur, India

Sonali Dhakate
UG Student, Department of Electronics and Telecommunications, Priyadarshini College of Engineering, Nagpur, India
dhakate.sona@gmail.com

Mrunal Shegaonkar
UG Student, Department of Electronics and Telecommunications, Priyadarshini College of Engineering, Nagpur, India
shegaonkarmrunal@gmail.com

Payal Sahare
UG Student, Department of Electronics and Telecommunications, Priyadarshini College of Engineering, Nagpur, India
payalsahare20@gmail.com

Khushi Ukey
UG Student, Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
khushiukey5678@gmail.com

***Abstract* - The management of airlines depends on the forecasting of air passenger flow, but standard forecasting techniques cannot guarantee the accuracy of the forecast. When they encounter large-scale, multidimensional, nonlinear, and non-normal distributing time series data, they have the ability to generalize. In this paper the SVM regression is implemented to help with forecasting air passenger flow. We discover that the SVM regression algorithm's outcome exhibits the least inaccuracy when compared to the other two forecasting techniques by carefully choosing the parameters and kernel function. Concerns about demand splits among service providers have been given to every sector by the nearperfect competition scenario. This is especially important in the airline business, where high service standards are the norm. Demand is driven by the player who achieves the greatest mappings between every one of his the airline's offerings and the set of consumer preferences. The airline business has grown exponentially as a result of the economic reforms of 1991, which were promptly followed by the privatization of Indian skies, creating nearly ideal competition. More specifically, cross-border activities have started in the international sectors, where previously only domestic carriers operated.**

***Keywords* – Airline Passengers, Support Vector Machine, Forecasting, Machine Learning**

## I. INTRODUCTION

Forecasting the flow of air passengers is a crucial step in the administration of an airline, one that involves both operations management and the fleet assignment program commonly known as FAP. Many academics have studied it since an accurate estimate of air passenger traffic is crucial to managing an airline's profitability. There are some established forecasting methods. Time series analysis, logistic regression analysis, the grey theory, combination forecasting, and other techniques [1] are a few examples. The research in the majority of the aforementioned techniques focuses on the time a sequence model and the relation regression model. They could provide us with an approximation of the passenger following trend. However, when dealing with large-scale, multi-dimensional, complex, and data with nonnormal distribution. It cannot ensure that a model can be generalized. The shortcomings of traditional methods have been somewhat mitigated by the artificial intelligence (AI) algorithm of neural networks, which not only has the capacity for generalization and nonlinear mapping but also for strong robustness and better forecast accuracy.

When the sample size is small, it is simple for this technique to reach an agreement on local minima because it largely relies on the experience risk minimum principle [2].

The aforementioned issue has been effectively resolved by the SVM regression model. The SVM algorithm, a new algorithm built on the structure risk minimal principle, is superior to other algorithms built on the experience risk minimum principle. Theoretically, the globally optimal solution can be obtained by applying support vector machines to the function regression estimation from the SVM regression problem, which is believed to be a convex quadratic programming problem [3] [4].

## II. LITERATURE SURVEY

The research focused on the routes from Jakarta to Yogyakarta and Jakarta to Singapore, which are examples of the most lucrative routes for domestic as well as global flights. The study used regular passenger information provided by Indonesian airlines. Subsequent to a comparison of the methods' Mean Absolute Percentage Errors (MAPE), predicted future demand for the following 12 months was determined. With MAPE, or mean of 1.29 for the CGK-JOG journey and 2.2 for the CGKSIN route, neural networks outperformed ARIMA in both routes. [5]

The demand for seats is essentially unpredictable, the capacity is fixed and hard to expand, and the varying expenses are extremely expensive. The data employed for this research are from a prominent Turkish airline and cover the previous five years' Worth of daily flight passenger data. For 355 days that the flight is available for purchase, it is used to predict the anticipated passenger load in reservation systems. For predicting, two methods—Box Jenkins and artificial neural networks—are used. These approaches are contrasted in the end. [6]

The authors will analyze the passenger load from a prior pattern in this paper and develop a model of estimation using decision trees in order to forecast the traveler load based on specified parameters. The root indicates that when the predicted outcome is tested in contrast to the real data given for a

particular month, the square variation is observed for all of the airlines operating at the airport. It shows the model's value in predicting actual passenger traffic, which is crucial for managing employees at runways for day-to-day operations.
Finally, using the expected volume of individuals as input, a simulation model has been developed that determines the precise number of check-in counters required to satisfy all requirements of the service level agreement. [7]

In this study, we analyze passenger load from past periods of pattern in order to forecast the passenger weight based on particular parameters and develop a prediction model using decision trees. When the model is evaluated in contrast to the true information given for a particular month, the root indicates that a rectangular variation is observed for all of the airlines operating at the airport. It shows how valuable the model is for forecasting actual passenger traffic, which is crucial for managing the use of resources at airlines for day-to-day operations. In the end, a simulation model has been developed using the expected passenger traffic as input, which determines the precise number of check-in counters required to satisfy the conditions of the service level commitment. [8]

The authors gathered data for this research on flight arrival times, on-time performance, and wait times from 48 foreign hub airports. Additionally, based on factors like an airport's performance in terms of flights heading on time, the number of flights, ranking for moving away on time, typical delays, and queue length, we used the Logistics Model Trees ML technique to predict the level of passenger satisfaction. The results are shared with others for deeper investigation and comprehension. [9]

Additionally, this study employs a Bidirectional LSTM model with a test set accuracy of 95.27% to classify three different types of emotion based on more than 15,000 tweets. These tweets show significant elements that led to negative reviews, such as subpar flight service, canceled or delayed flights, inconvenient bookings, and lost or damaged luggage. Aerospace companies ought to concentrate on solving these problems. As a result, the bilateral LSTM model could precisely identify the emotional inclinations of tweets, assisting aerospace companies in quickly changing the products and services they provide and better meeting consumer demand. [10]

The number of travelers and the price at which they pay, which differ for each flight, determine the revenue. The effects of social, political, and economic shifts on aviation are very likely. As a result, consumer behavior among travelers is quite active. Therefore, it is difficult to create a technique that will accurately project the earnings for each route. We will employ a semi-supervised learning method to get around this. Along with the current income from tickets, we also have the current list of passengers. Everyone also have information on prior passengers who purchased tickets. The majority of the knowledge is thus readily accessible, but one unknowable factor that may have significant effects on traveler travel patterns is shifting market conditions. [11]

Projected levels of air passenger flow are crucial for managing airlines, but when confronted alongside large-scale, multi-dimensional, nonlinear, and unusual transportation time series data, typical forecasting techniques fall short. This paper introduces the SVM projection algorithm in an effort to improve the forecasting of air passenger numbers. By carefully selecting each parameter and kernel function, the authors' find that the SVM classification algorithm's output displays the smallest amount of error when compared with the other two forecasting methods. [12]

This study suggests an innovative approach to forecasting high-speed train passenger volumes in a cutthroat market based on the XGBoost model. The technique also accounts for the cost of the train ticket. First, statistical analysis features are calculated to supplement the absent data and preprocess the flight booking information data. Then, the information about train and airline reservations is combined into a sizable training dataset, and a prediction model is built using the XGBoost. By using this technique for prediction, it is possible to determine passenger movement in a cutthroat market. According to experimental findings, the suggested model is more accurate at making predictions than conventional models. [13]

## III. METHODOLOGY

The suggested method for predicting airline passenger numbers is presented in this section. It comprises a sequential layer, two LSTM layers, and a dense layer. The suggested network's architecture is depicted in Figure 1.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

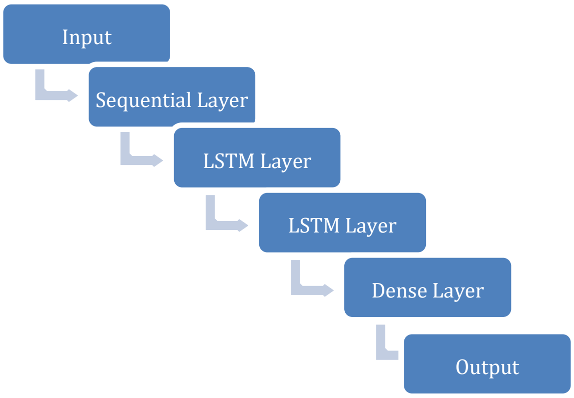



Figure 3.1 Proposed Structure

**Sequential Layer:**
Predicting a category for a sequence of inputs that spans space or time is the goal of the predictive modeling issue known as sequence classification. For a simple stack of segments each layer has precisely one input tensor alongside a single output tensor, a sequential model is suitable.

**LSTM Layer:**
In time series and sequences of data, an LSTM layer—a type of RNN—learns long-term relationships between time increments. The layer conducts additive interactions, which during training can enhance gradient flow over lengthy sequences.

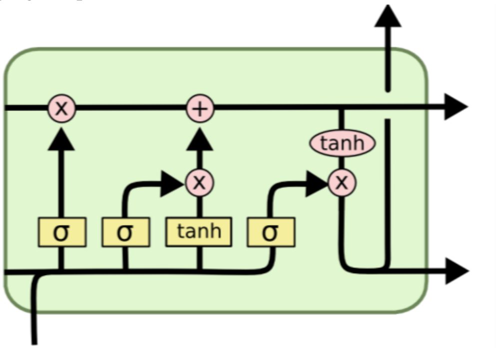



Figure 3.2 LSTM Model [14]

**Dense Layer:**
In any sort of neural network, a layer that is densely connected to its preceding layer implies that every neuron in the layer is connected to every other neuron in the layer above it. In artificial neural networks, this layer is the one that is most frequently used. [15]

**Implementation Process:**
The steps are clearly explained for implementing the procedural structure.

Step 1: The dataset collected from "airline_passenger_csv" and pandas read_csv() function was used for loading the dataset.

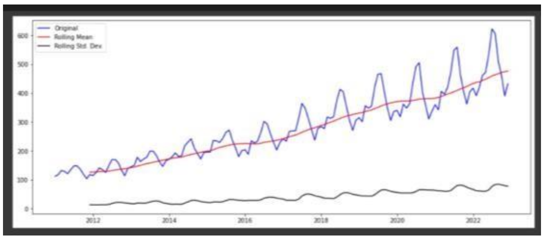
Step 2: sklearn's function MinMaxscaler() was used to train, test and normalize the data.

Step 3: The normalized data then divided into two parts, one for training that is of 65% and another for testing that is of 35% for testing.

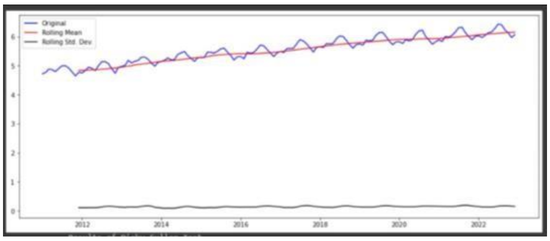
Step4: Then the divided data was pre-processed and the size of file were reduced to feed into the model.

Step 5: The model was trained successfully and results were obtained.

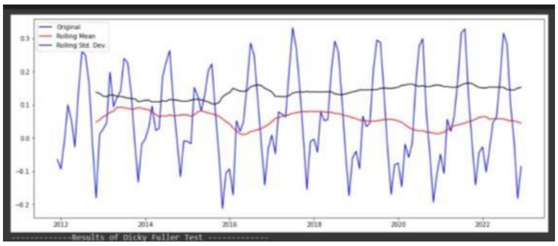
## IV. RESULT



4.1 Result Image 1

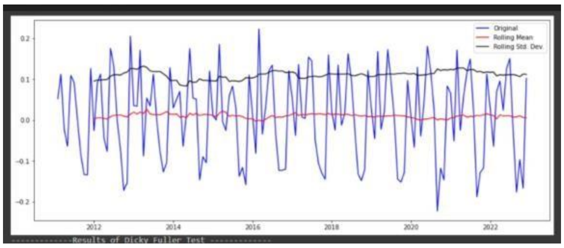


4.2 Result Image 2

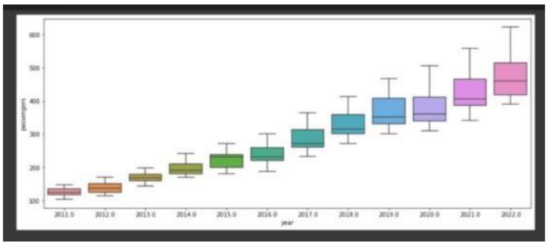


4.3 Result Image 3

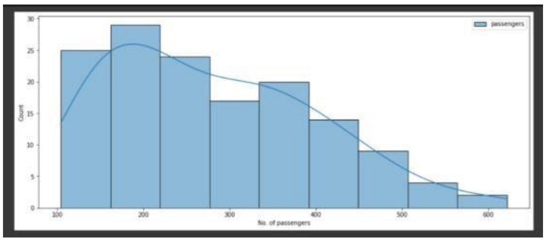
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930



4.4 Result Image 4



4.5 Result Image 5



4.6 Result Image 6

## IV. CONCLUSION

The forecasting of aircraft passengers is essential for airline investors and promoters when making strategic decisions. In this paper, we suggest an LSTM-based recurrent neural network-based system for forecasting the number of airline passengers. The proposed scheme works better than the other existing schemes, according to the results. By including convolution layers in the model's structure and attempting to make the model even deeper using additional LSTM layers, the suggested work can be further extended in the future.

## REFERENCES

[1] Zhai Yan-juan, "“The study on airline’s FAP in hub-and-spoke network”,," *Chinese commercial aviation university.*, pp. pp.18-26, (2016).

[2] X. Guo-en, ",Model of customer churn prediction on support vector machine,," *Systems Engineering-Theory & Practice,*, vol. vol.01, pp. pp.71-77, 2008,.

[3] S. .. Gunn, "“Support vector machines for classification and regression”,," *University of Southampton,*, pp. pp.152-156 , 1998.

[4] J. V. Suykens JAK, "Least squares support vectors machine classifiers,," *Neural processing Letters,*, pp. pp.293-300,
1999,9(3).

[5] A. D. a. E. L. S. Ramadhani, ""Airline Passenger Forecasting using ARIMA and Artificial Neural Networks Approaches,"," *2020 International Conference on ICT for Smart Society (ICISS),*, pp. pp. 1-5,, Bandung, Indonesia, 2020, .

[6] S. M. T. F. Ghomi and K. Forghani, ""Airline passenger forecasting using neural networks and Box-Jenkins,"," *2016 12th International Conference on Industrial Engineering (ICIE),*, pp. pp. 10-13,, Tehran, Iran, 2016,.

[7] M. C. a. P. S. M. N. Laik, ""Predicting Airline Passenger Load: A Case Study,"," *2014 IEEE 16th Conference on Business Informatics,*, pp. pp. 33-38, Geneva, Switzerland, 2014, .

[8] B. V. V. a. V. R. H. B. Sankaranarayanan, ""An exploratory analysis for predicting passenger satisfaction at global hub airports using logistic model trees,"," *2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN),*, pp. pp. 285-290,, Kolkata, India, 2016,.

[9] B. V. V. a. V. R. H. B. Sankaranarayanan, ""An exploratory analysis for predicting passenger satisfaction at global hub airports using logistic model trees,"," *2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN),* , pp. pp. 285-290,, Kolkata, India, 2016,.

[10] C. Tan, ""Bidirectional LSTM Model in Predicting Satisfaction Level of Passengers on Airline Service,"," *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE),* , , no. doi: 10.1109/ICAICE54393.2021.00107., pp. pp. 525-531,, Hangzhou, China, 2021.

[11] A. N. M. R. S. P. l. K. T. W. M. A. l. P. D. P. N. a. D. K. P. H. K Tissera, ""Predictive Analytics Platform for Airline Industry,"," *2020 2nd International Conference on Advancements in Computing (ICAC),*, no. doi: 10.1109/ICAC51239.2020.9357244., pp. pp. 108-113,, Malabe, Sri Lanka, 2020,.

[12] Y. Ke-wu, ""Study on the Forecast of Air Passenger Flow Based on SVM Regression Algorithm,"," *2009 First International Workshop on Database Technology and*

*Applications*,, no. doi: 10.1109/DBTA.2009.33., pp. pp. 325-328,, Wuhan, China, 2009.

[13] S. M. T. F. Ghomi and K. Forghani, ""Airline passenger forecasting using neural networks and Box-Jenkins,"," *2016 12th International Conference on Industrial Engineering (ICIE),* , no. doi: 10.1109/INDUSENG.2016.7519342., pp. pp. 10-13,, Tehran, Iran, 2016,.

[14] Z. J. M. G. W. H. a. S. X. H. Huiting, ""Research on Forecast of Passenger Flow of High Speed Railway in Competitive Market Based on XGBoost Model,"," *2020 13th International Symposium on Computational Intelligence and Design (ISCID),* , no. doi: 10.1109/ISCID51228.2020.00031., pp. pp. 110-113,, Hangzhou, China, 2020,.