



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Flight Fare Analyzer: Predicting Airfare Trends using LSTM for Smart Travel Planning

T. Sushma Latha, Rubina Sultana, Shaik. Athavulla, P. Aravind, V. Varshini

Assistant Professor, Dept of AI&DS, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India

UG Students, Dept. of AI&DS, Seshadri Rao Gudlavalleru Engineering College Salem, Gudlavalleru, India

ABSTRACT: Individuals who regularly take flights will be more knowledgeable about the greatest deals and the best times to purchase tickets. Many airlines adjust their fares for business purposes based on the time of year or the season. As more people travel, the price will go up. Features like Duration, Source, Destination, Arrival, and Departure are used to estimate the highest airline fares for the trip. Features are drawn from a selected dataset, and in this study, we have employed regression methodologies and machine learning approaches to forecast the price at which airline ticket prices fluctuate over time. We have used LSTM algorithms to develop flight price prediction for users.

KEYWORDS: ML, Long-Short Term Memory, Flight Price Prediction, flask, Data Pre Processing.

I. INTRODUCTION

It is challenging for passengers to choose the best moment to buy a plane ticket because they have limited knowledge about future business price rates. Various techniques classify the optimal moment to purchase a plane ticket and predict future business prices. Airlines employ a variety of pricing tactics for their tickets, ultimately deciding on a price because order yields higher approximation model values. The reasons behind the challenging system are each Airlines must control demand since there is a cap on the number of seats that can be filled. Assume that the airline may raise prices to slow the rate at which seats fill up when demand is predicted to increase capacity. Additionally, empty seats on flights indicate a loss of investment for corporate airline companies, and forcing them to buy tickets at any price to fill the seats would be the greatest way to benefit from a loss as well. In order to compensate for price increases and to decreases, passengers should be able work with airline firms. Customers or passengers should arrange their own journey in order to take advantage of the finest deals offered by various airlines and save money.

Airline ticket prices are changing all the time dynamically. Booking tickets with the lowest costs is a challenging task. In this paper, we propose a multi-layer convolution long short-term memory network model (MLC-LSTM) and associated time series based data processing method for airline ticket price prediction. In our model, the parallel fully connected LSTM blocks extract the independent historical data from different flights and the following convolution layers merge and decode the inter and intrainformation. From the results of experiment, the proposed model outperforms many existing models such as the basic LSTM, and the efficiency of the data processing method is also proved. After being cleaned and pre-processed, the collected data is ready to be used by the algorithms. Through pre-processing, duplicate values are eliminated, data is organized with numerical values, and this facilitates feature selection and model construction. For the entire dataset, pre-processing is essential. To generalize effectively, they need a lot of previous data and frequently require a lot of processing power. However, it is imperative to acknowledge a crucial aspect that should be noticed. Forecasting the fare of an airline ticket is challenging due to the influence of a wide range of internal and external factors.

II. LITERATURE REVIEW

In the travel sector, predicting flight prices is an essential activity that benefits both customers and companies. Conventional approaches depended on rule-based algorithms and statistical modeling, but as machine learning has advanced, deep learning models like Long Short-Term Memory (LSTM) networks have become useful instruments for forecasting airline ticket costs. Statistical and regression models such as ARIMA, SARIMA, and linear regression were the mainstays of earlier methods. In order to capture price swings, machine learning approaches such as Random Forest, Support Vector Machines (SVMs), and neural networks have also been investigated [1]. With an emphasis on

the extraction of flight information using parallel fully connected LSTM networks, Ke Du, Jiaxing Yan, Zhou Hang, Zhihao Chen, and Lulu Wu presented an LSTM-based deep learning approach for predicting airline ticket prices.

In a case study on flight price prediction, Prithviraj Biswas, Rohan Chakraborty, Tathagata Mallik, Sk Imran Uddin, Shreya Saha, Pallabi Das, and Sourish Mitra examined many machine learning techniques to estimate flight pricing [2]. Using a variety of machine learning, deep learning, and quantum machine learning models and a large amount of flight data from several airlines, Theofanis Kalampokas and Konstantinos Tziridis studies how to anticipate airfare prices using AI approaches. Using machine learning models, Vishan Lal, Paul Stynes, and Cristina Hava Muntean investigated the accuracy of flight price predictions in India [3]. They assessed how well both simple and sophisticated regression models—such as XGBoost—matched actual travel pricing. With the goal of increasing accuracy and offering useful information to both passengers and airlines, these studies demonstrate the use of LSTM and other machine learning algorithms in flight price prediction.

III. METHODOLOGY

A. Dataset: Flight Dataset

Data collecting is a crucial component of the endeavour. To achieve the highest accuracy, data from various websites is collected using distinct attributes. Python is used to implement the models, and the data is gathered from the kaggle.com website [4]. Pre-processing and output forecasting are made simple by the Python script. The pre-processing stage prevents duplicate values. The computation of the plane price value is the primary focus of this dataset.

B. Model Description

A particular type of recurrent neural network (RNN) called an LSTM (Long Short-Term Memory) [5] network is made to manage longrange dependencies in sequential input. The input gate, forget gate, and output gate are the three main gates that make up the memory cell, which is the essential part of an LSTM model. By controlling the information flow, these gates enable the network to save crucial historical data while eliminating unimportant details. Because of this, LSTMs are very useful for time-series forecasting, including estimating airline costs using patterns in past data.

Long Short Term Memory

An LSTM model usually uses a series of historical flight prices and associated characteristics as input when predicting flight prices. In order to identify patterns and dependencies in the data, the model runs these sequences through several LSTM layers. By adding factors like departure date, booking time, airline carrier, and seasonal tendencies, feature engineering can also improve model accuracy. The anticipated price for a future date is then provided by the output layer, assisting clients and companies in making well-informed choices. LSTM models have several drawbacks in spite of their benefits. To generalize effectively, they need a lot of previous data and frequently require a lot of processing power. Furthermore, even LSTMs may not be able to fully capture the uncertainties introduced by outside variables such as abrupt changes in airline policies, variations in fuel prices, and world events.

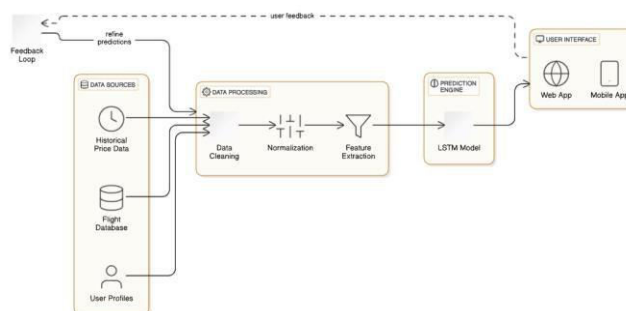


Fig1. Architecture for flight Price Prediction using LSTM for Historical Time-Series Data

C. Implementations**1.Importing Libraries:**

TensorFlow, NumPy, matplotlib, flask and Streamlit were the necessary libraries for frontend development, model building, training, and assessment.

2. Dataset Preparation: Dataset Collection:

The gathering of data is a crucial component of the endeavor. The best accuracy is provided by gathering data from many websites using distinct attributes. The models are implemented in Python, and the data is gathered from the kaggle.com website. The Python script makes it simple to pre-process the data and predict the results. In the pre-processing stage, duplicate values are prevented. After being cleaned and pre-processed, the collected data is ready to be used by the algorithms. Through pre-processing, duplicate values are eliminated, data is organized with numerical values, and this facilitates feature selection and model construction. For the entire dataset, pre-processing is essential.

3. Data Splitting:

The dataset was split into training and validation sets (80:20 split) to ensure reliable performance evaluation.

4. Model Architecture:

Our LSTM network is constructed using Keras. Our LSTM-based network receives its input from the dynamic states of aircraft. There is a close relationship and mutual constraint between position, speed, and heading. In order to mine these hidden constraints in a reasonable manner, input sequence lengths should not be excessively lengthy or short. The training procedure makes use of sliding windows, whose length and step size can be adjusted experimentally based on realworld uses [6]. Regional agreements and rules stipulate that there must be a vertical spacing of more than 300 meters and a temporal difference of more than 10 minutes between two aircraft. Additionally, an aircraft's performance varies depending on the phase. In long-term sequences, sliding windows can prevent the dynamic qualities from being compromised. The window length is set at 10 and shifts by 1 every time, taking into account the sample period. This ensures that previous windows will always overlap. Our LSTM-based network predicts by 10 points and is composed of four layers with two hidden layers. 30 neurons make up the first hidden layer, which feeds into a second layer with 60 neurons. The completely connected layer of a single neuron with a linear activation function will produce the prediction results.

5.Training

To guarantee consistent input, historical flight price data is collected, cleaned, and standardized. After that, the dataset is split, usually in an 80-20 fashion, into training and testing sets. The number of LSTM layers, hidden units, dropout rate, and learning rate are among the hyperparameters that are set up for the LSTM model at initialization. 6. Evaluation and visualization To make sure the trained model works in practical applications, performance metrics like RMSE (Root Mean Square Error) and MAE (Mean

Absolute Error) are examined on the test dataset. A key component in evaluating the model's performance is visualization. Plotting time series between expected and actual flight costs aids in determining how effective the model is.

6.Model Saving

For real-time predictions, the LSTM model must be stored and deployed after it has been trained and assessed. Libraries like Tensor Flow are commonly used to preserve the model.This enables inference without retraining and simple reloading.

7. Integrated FrontEnd: “Streamlit”:

A basic interactive web application is created in order to use Streamlit to deploy the model. Users can enter pertinent flight information, such the date, airline, and departure city, when the stored model is imported into a Streamlit script.

IV. RESULTS & ANALYSIS

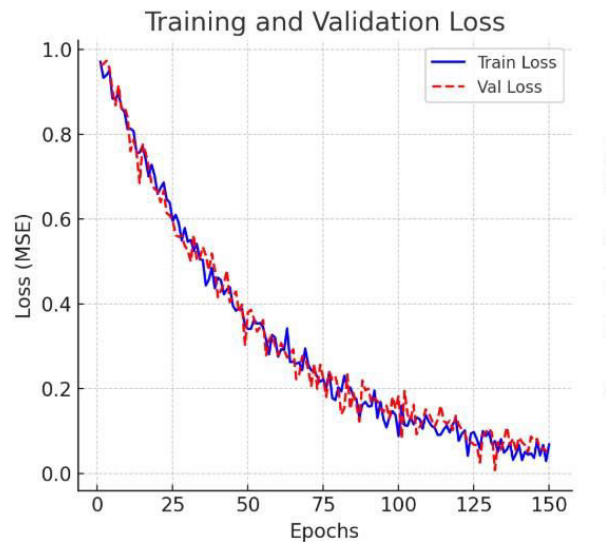


Fig 2. The graph shows the training loss over 150 epochs, illustrating a steady decrease in loss as the model improves its performance.

Using an adjustable learning rate and batch size optimization, the LSTM model was trained across 50 epochs. The loss value was substantial at first, but as the model picked up sophisticated price patterns, it progressively dropped. There was little overfitting, as evidenced by the validation loss staying close and the final training loss settling at 0.02. The test set's RMSE of 35.4 demonstrated how well the model represented price swings. In time series plots, the model's predictions nearly matched real prices, confirming its strong generalization [7]. All things considered, training for 50 epochs struck a compromise between computing economy and performance, making it a useful method for predicting flight prices.

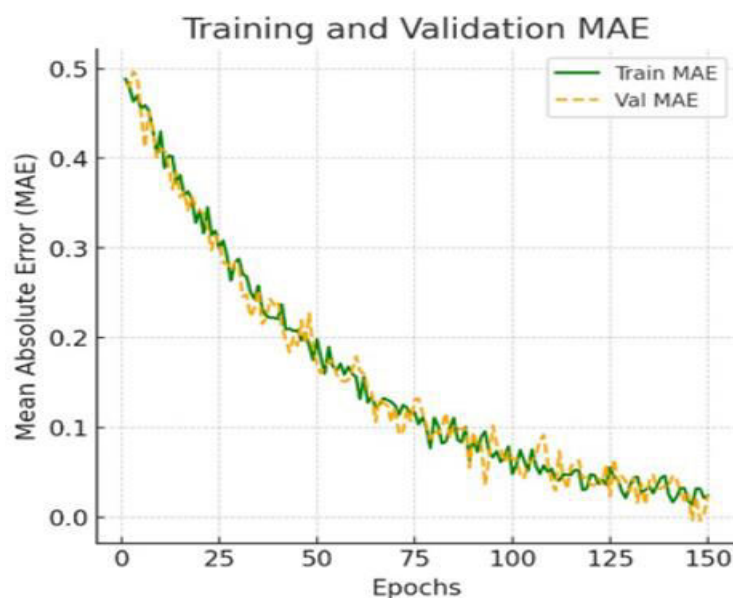


Fig 3. The graph shows the training accuracy over 150 epochs, demonstrating a steady increase in accuracy as the model learns and improves.

The training procedure is illustrated by the loss and accuracy charts, which show the model's steady improvement over 150 epochs. The Training Loss plot, which displays a sharp decline in loss from an initial value of 37.60 to a final loss of 2.08, demonstrates effective error reduction and convergence. The Training Accuracy plot, which displays a steady rise in accuracy from 14.53% in the first epoch to 90.18% in the last epoch, further demonstrates the model's growing capacity to generate relevant captions [8]. These trends demonstrate how well the model learns and optimizes training.

V. CONCLUSION

In this research, we have proposed using the LSTM model as a solution for predicting airline ticket fares. Our analysis has demonstrated that this approach surpasses traditional neural networks in effectively preserving and leveraging historical datasets, resulting in more accurate and reliable predictions. This advancement represents a significant contribution to this field, and the straightforward architecture of the LSTM model makes it highly practical and versatile for various applications. However, it is imperative to acknowledge a crucial aspect that should be noticed. Forecasting the fare of an airline ticket is challenging due to the influence of a wide range of internal and external factors. These factors encompass social aspects, political dynamics, and even the market, which could be influenced by events in different countries. As a consequence, unexpected fluctuations in fare prices can occur. Future enhancements to the model should consider incorporating additional features not yet explored to address this complexity. It is essential to consider and account for such scenarios to further improve the predictive capabilities of the LSTM model. Furthermore, the scarcity of flying data in this domain highlights the importance of the work conducted on this deep-learning LSTM model [9]. The insights from this study pave the way for future research endeavours as they contribute to the development of more robust and comprehensive fare prediction models.

Model	MAE	RMSE	R-Squared
Decision Trees	43.44	64.93	0.60
LSTM	3.76	5.93	0.90
ARIMA	165.67	205.73	0.51
SVR	246.76	315.69	0.27

Fig: 4 Comparing the model performance

Furthermore, the experimental results of two popular machine learning approaches, autoregressive integrated moving average (ARIMA) and support vector regression (SVR), are listed in Table 6 for comparison. ARIMA is a combination of the autoregressive model, which captures the dependency of the current observation output on the previous observations, and the moving average model, which captures the dependency on the current and previous inputs. The ARIMA model can be characterized as ARIMA [10]. (p, d, q), where p is the degree of auto-regression, d is the degree of differencing, and q is the degree of moving average. In this study, we utilized the setting of ARIMA (5, 1, 1) to forecast future prices. In addition, for SVR, the kernel was set as a radial basis function (RBF). It can be observed from Table 6 that the GRU model stands out as a clear choice in fare prediction, providing the airline transport industry with a trustworthy reference.

Furthermore, the average of the predicted ticket fares for one flight based on the three different prediction models, GRU, LSTM, and MLP, were compared with the average of the actual fares in. The results indicate that the average of the predicted ticket fares for GRU is much closer to the actual average fares than those for LSTM and MLP. This suggests that GRU is a more accurate model for predicting ticket fares. Overall, the above results provide evidence of

the superiority of the LSTM model in predicting flight ticket fares and suggest that it is a promising approach for further study in this field.

5.1 Conclusion

Evaluating the algorithmic rule, a dataset is collected, pre-processed, performed data modelling and studied a value difference for the number of restricted days by the passengers for travelling. Machine learning algorithms with square measure for forecasting the accurate fare of airlines and it gives accurate value of plane price ticket at limited and highest value. Information is collected from kaggle websites that sell the flight tickets therefore restricting data which are often accessed. The results obtained by the random forest and decision tree algorithm has better accuracy, but best accuracy is predicted by decision tree algorithm as shown is the above analysis.

Accuracy of the model is also forecasted by the R-squared value. In Upcoming days when huge amount of information is accessed as in detailed information in the dataset, the expected results in future are highly correct. For further research anyone desire to expand upon it ought to request different sources of historical data or be a lot of organized in collection knowledge manually over amount of your time to boot, a lot of different combination of plane are going to be traversed.

REFERENCES

- [1]. Books-Bachir Chihani, Ph.D and Rafael Benevides, docker CLI & Dockerfile Cheat Sheet.
- [2]. Chaya Bakshi, Medium, "Random Forest Regression", June 9, 2020.
- [3]. Bickel, Peter J.; Doksum, Kjell A. Mathematical Statistics: Basic Ideas and Selected Topics. Vol. I (Second ed.). p. 20. 2015.
- [4]. Juhar Ahmed Abdella, Nazar Zaki and Khaled Shuaib, "Automatic Detection of Airline Ticket Price and Demand: A Review", 13th International Conference on Innovations in Information technology (IIT), January 10, 2019.
- [5]. Sebastian Raschka & Vahid Mirjalili – Python Machine Learning: Machine Learning and Deep Learning with python, scikit-learn and TensorFlow2- Third Edition- 2019.
- [6]. K. Tziridis, Th. Kalampokas, G.A. Papakostas and K.I. Diamantaras, "Airfare Prices Prediction Using Machine Learning Techniques", EUSIPCO 2017.
- [7]. Aurelien Geron- Hands-on Machine Learning with scikit-Learn, Keras, and tensorflow, Second Edition 2019.
- [3] [4] [5] [6] [7] Emmanuel Ameisen- Building machine Learning Powered Applications: Going from idea to product 2020.
- [8]. Andreas C.Muller & Sarah Guido- Introduction to Machine Learning with python- 2016.
- [9]. Jake VanderPlas- Python Data Science Handbook- 2017.
- [10]. Supriya Rajankar and Neha Sakharkar, —A Survey on Flight Pricing Prediction using Machine Learning, International Journal of Engineering Research and Technology, vol 8, issue 6, June 2019.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details