
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

AIRLINE FARE PREDICTION USING MACHINE LEARNING ALGORITHMS

[1]Dr M Sandhya Rani, [2]Sri Charitha Kalakonda, [3]Srilekha Kambala

[1]Associate professor in Department of Information Technology Bhoj Reddy Engineering College for Women

[2,3] UG Scholars in Department of Information Technology Bhoj Reddy Engineering College for Women

Abstract

Optimal timing for airline ticket purchasing from the consumer's perspective is challenging principally because buyers have insufficient information for reasoning about future price movements. This paper deals with the problem of airfare prices prediction. For this purpose a set of features characterizing a typical flight is decided, supposing that these features affect the price of an air ticket. The features are applied to eight state of the art machine learning (ML) models, used to predict the air tickets prices, and the performance of the models is compared to each other. Along with the prediction accuracy of each model, this paper studies the dependency of the accuracy on the feature set used to represent an airfare.

Keywords: Airfare, Machine Learning algorithms, Predictions, Flight, Linear Regression

INTRODUCTION

According to a report, India's civil aviation industry is on a high-growth trajectory. India aims to become the third-largest aviation market by 2020 and the largest by 2030. Indian domestic air traffic is expected to cross 100 million passengers by FY2017, compared to 81 million passengers in 2015, as per Centre for Asia Pacific Aviation (CAPA).According to Google Trends, the search term - "Cheap Air Tickets" is most searched in India. Moreover, as the middle-class of India is exposed to air travel, consumers hunting for cheap prices increases. Machine Learning is one of the most hot research topics in computer science and engineering, which is applicable in many disciplines. It provides a collection of algorithms, methods and tools able to embody some kind of intelligence to machines. The power of ML is the provided modeling tools, which are able to be trained, via a learning procedure; One of the reasons that ML has attracted scientists from several disciplines is its ability to provide human-like intelligence to machines as the amount of data used during learning increases. However, the increase of the training data needs parallel implementations of the ML algorithms using specialized software and/or hardware platforms. In the context of machine learning, there are two possible alternatives for handling the problem of airfare pricing prediction. The first approach tackles the prediction of air tickets prices as a regression problem, while the second one transforms it to a classification task. The former strategy is usually applied for the prediction of the exact air ticket price, since the regression models try to approximate a function that describes the mapping law between data features and airfare prices. The later approach cannot predict the exact air ticket prices, but can provide decisions regarding the range of a price or a decision to buy or not the ticket with the specific price. One of the reasons that ML has attracted scientists from several disciplines is its ability to provide human-like intelligence to machines as the amount of data used during learning increases. However, the increase of the training data needs parallel implementations of the ML algorithms using specialized software and/or hardware platforms. In the context of machine learning, there are two possible alternatives for handling the problem of airfare pricing prediction. The first approach tackles the prediction of air tickets prices as a regression problem, while the second one transforms it to a classification task. The former strategy is usually applied for the prediction of the System

LITERATURE SURVEY

WILDTRACK AI-ANIMAL CLASSIFICATION RELATED WORK:

Description: The study begins by collecting and preprocessing a large dataset of flight information, including attributes such as departure and arrival times, durations, airlines, and prices. Feature engineering techniques are then applied to extract meaningful features from this data, such as the day of the week, time of year, and route popularity. Next, the study trains and evaluates four different machine learning models on the preprocessed data: decision trees, support vector machines, k-nearest neighbors, and random forests. Each model is assessed using metrics such as mean squared error (MSE) and R-squared, which measure the difference between predicted and actual prices, as well as the proportion of variance explained by the model.
M. Papadakis, “Predicting Airfare Prices,” 2014[2]
Description: This paper presents an innovative approach to flight fare prediction by combining multiple machine learning algorithms to form an ensemble model. The ensemble consists of decision trees, gradient boosting, and linear regression, each contributing their unique strengths to improve prediction accuracy. To further enhance the model's performance, feature selection techniques are employed to identify the most relevant features from the dataset. The ensemble model is evaluated using metrics

IJMECE
International Journal of modern electronics and communication engineering
ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), which assess the model's ability to accurately predict flight fares. The results demonstrate that the ensemble approach outperforms individual models, achieving a significant reduction in RMSE and MAPE.

"Gordiievych and I. Shubin, "Forecasting of airfare prices using time series,"

The study finds that ARIMAX outperforms the other two models, achieving the lowest AIC values and highest forecasting accuracy. This suggests that incorporating exogenous variables, such as economic indicators and weather data, improves the predictive power of the model. STL decomposition also performs well, effectively capturing seasonal patterns in flight fares. In contrast, ARIMA model struggle to account for these patterns, resulting in lower accuracy. This research provides valuable insights for the airline industry, highlighting the importance of selecting appropriate time series analysis techniques for flight fare prediction. By identifying the most effective approach, airlines can optimize their pricing strategies and maximize revenue. The study's findings also contribute to the broader field of time series analysis, informing the development of more accurate predictive models.

"***Prices, prices and prices: A study in the airline sector." Tourism Management***,

Description: The study assesses the performance of the hybrid model using metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The results demonstrate a significant reduction in RMSE and MAE compared to individual models, indicating the effectiveness of the hybrid approach. The study's findings have important implications for the airline industry, as accurate flight fare prediction can lead to optimized pricing strategies and increased revenue. The proposed hybrid model offers a robust and adaptable solution for flight fare prediction, and its application can be extended to other domains where time series forecasting is crucial

EXISTING SYSTEM

Flight ticket prices can be something hard to guess, today we might see a price, check out the price of the same flight tomorrow, and it will be a different story. We might have often heard travelers saying that flight ticket prices are so unpredictable. As data scientists, we are gonna prove that given the right data anything can be predicted.

Problems in Existing System:

• ***Data Dependency***::ML algorithms require large amounts of data to make accurate predictions. In the case of airline fares, historical pricing data, weather

****Complexity***: Predicting airline fares involves dealing with a multitude of variables, such as time of booking, route popularity, seasonality, and competitor pricing. Managing the complexity of these factors within ML models can be challenging.

• ***Dynamic Nature***::Airline fares are highly dynamic and can change rapidly due to various factors such as demand, fuel prices, and competitor actions. ML models need to be continuously updated and retrained to adapt to these changes effectively.

•***Accuracy:*** Achieving high accuracy in fare prediction can be difficult due to the volatility of the market and the influence of unpredictable events such as natural disasters, geopolitical tensions, or regulatory changes.

• ***Over fitting***: ML models may over fit the training data, resulting in poor generalization to new data. Balancing model complexity and generalization ability is crucial in ensuring accurate predictions

PROPOSED SYSTEM

The goal in this step is to develop a benchmark model that serves us as a baseline, upon which we will measure the performance of a better and more tuned algorithm. We are using different Regression Technique and comparing them to see which algorithm is giving better performance than other and At the end we will combine all of them using Stacking and see how our model is predicting

Advantages

Able to provide the right time to the buyer to purchase an air ticket by predicting the airfare price.

• Travelers can plan their trips more effectively by getting insights into how fares are likely to change in the future.

• Improve customer service by providing travelers with more accurate and timely information about fares

IMPLEMENTATION

Gathering Data: Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems. In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices. It is one of the most important steps of the life cycle. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data,the more accurate will be the prediction.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps

Data preparation

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to

IJMECE
International Journal of modern electronics and communication engineering
ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

use in our machine learning training. In this step, first, we put all data together, and then randomize the ordering of data. This step can be further divided into two processes:

- Data exploration:

It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

- Data pre-processing

Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format.

It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

It is not necessary that data we have collected is always of our use as some of the data may not be useful. In real-world applications, collected data may have various issues, including:

- Missing Values
- Duplicate data
- Invalid data

So, we use various filtering techniques to clean the data.It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome AIRLINE FARE PREDICTION DESIGN Department of IT, BRECW Page 24 Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model. Hence, in this step, we take the data and use machine learning algorithms to build the model.

Train Model

Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem. We use datasets to train the model using various machine learning

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

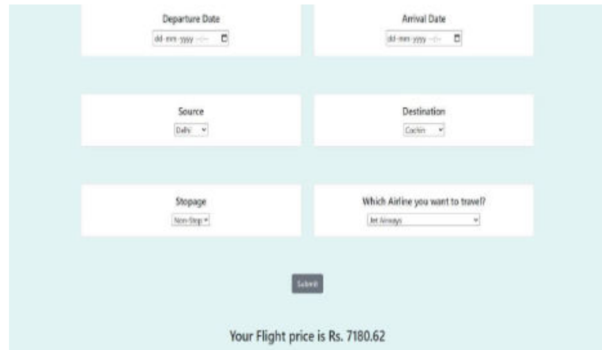
Test Model

Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it. Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

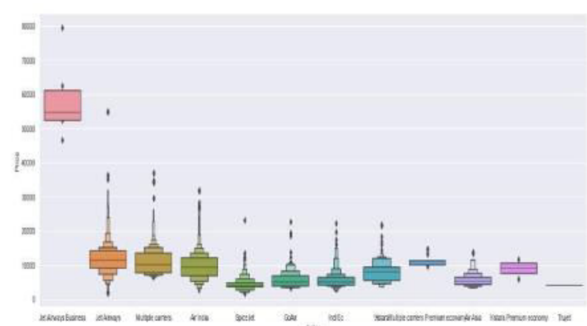
Deployment

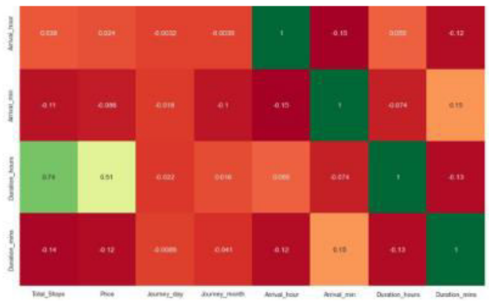
The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system. If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data

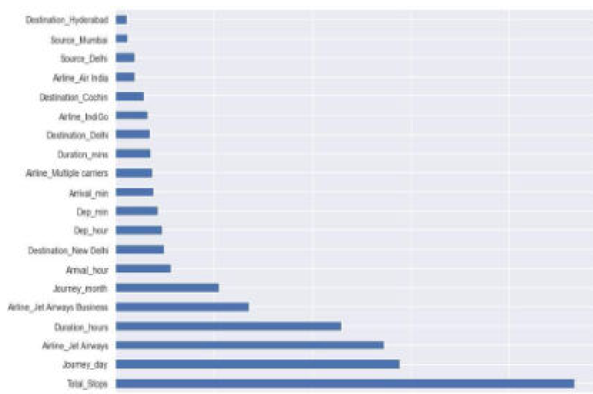
RESULTS



	Airline	Source	Destination	Route	Total_Stops	Additional_Info	Price	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min	Dur
0	IndiGo	Banglore	New Delhi	BLR → DEL	non-stop	No info	3897	24	3	22	20	1	10	
1	Air India	Kolkata	Banglore	CCU → IXR → BBI → BLR	2 stops	No info	7662	1	5	05	50	13	15	
2	Jet Airways	Delhi	Cochin	DEL → LKO → BOM → COK	2 stops	No info	13882	9	6	09	25	4	25	
3	IndiGo	Kolkata	Banglore	CCU → NAG → BLR	1 stop	No info	6218	12	5	18	05	23	30	
4	IndiGo	Banglore	New Delhi	BLR → NAG → DEL	1 stop	No info	13302	1	3	16	50	21	35	







CONCLUSION

This project deals with the problem of airfare prices prediction. For this purpose a set of features characterizing a typical flight is decided, supposing that these features affect the price of an air ticket. The features are applied to eight state of the art machine learning (ML) models, used to predict the air tickets prices, and the performance of the models is compared to each other. Along with the prediction accuracy of each model, this paper studies the dependency of the accuracy on the feature set used to represent an airfare. For the experiments a novel dataset consisting of 1814 data flights of the Aegean Airlines for a specific international destination (from Thessaloniki to Stuttgart) is constructed and used to train each ML model. The derived experimental results reveal that the ML models are able to handle this regression problem with almost 88% accuracy, for a certain type of flight features.

REFERENCES

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

[2] M. Papadakis, “Airline fare prediction using machine learning algorithms,” 2014

[3] Gordiievych and I. Shubin, " Airline fare prediction using machine learning algorithms," 2015 Information Technologies in Innovation Business Conference (ITIB), 2015, pp. 68-71, doi: 10.1109/ITIB.2015.7355055.

[4] Y. Narangajavana, F. J. Garrigos-Simon, J. S. García,S. Forgas-Coll, “Prices, prices and prices: A study in the airline sector.” Tourism Management, 41, 28-42, 2014

[5] Lantseva, K. Mukhina, A. Nikishova, S. Ivanov, and K. Knyazkov, “Data-driven

modeling of airlines pricing." Procedia Computer Science, 2015, 66, 267-276.

[6] H. C. Wen and H. P. Chen, "Passenger booking timing for low-cost airlines: A continuous logit approach." Journal of Air Transport Management, 64, 91-99, 2017.

[7] S. J. Dominguez-Menchero, J. Riveraand E. Torres-Manzanera, "Optimal purchase timing in the airline market." Journal of Air Transport Management, 40, 137143, 2014.