ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/261-263
Sanaboina Renuka Tejaswini / International Journal of Management Research & Review

Flight Price Prediction Using Machine Learning

Sanaboina Renuka Tejaswini
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
Ch.Jeevan Babu
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

ABSTRACT: ***Machine learning is a field of study that works with computer algorithms and data. It basically builds a tool on sample data and then predicts or take decisions about the same data without the need of specific programming. ML has its use in various industries including aviation industry where it is used to predict flight prices.***

In today's world, flights have become a common mode of transportation for humans. It is faster way to reach destinations, thus saving a lot of time. But also not everyone can afford a plane ticket and fly high in the sky. The ticket prices for a plane journey are very high when compared to that of a train or a bus. But irrespective of that, nowadays the number of people using flights have increased massively. Thus it becomes very hectic to maintain the prices of tickets with changing conditions. The airline companies use various complex techniques to predict the various flight prices using various factors present at that time. These factors commonly include market related issues, financial issues and various social issues. So it becomes difficult for them to do so. Also the consumer has no such tool in hand to have an idea about the ticket prices, that could very helpful.

This project is designed and developed keeping in mind this problem and develops an algorithm that predicts various flight prices keeping in mind various factors that affects them. This can help the airline companies to check what prices they could maintain. Also it would help the customer to predict future flight ticket prices, which would help in planning the trip accordingly.

I. Introduction

Machine learning is the part of artificial intelligence that provides a computer the ability to automatically learn with experience without the need of external programming. It is one of the hot ,most in demand research topic in computer science engineering. It can provide intelligence to the machines with the help of various tools and techniques. ML uses programs that can access data and learn for themselves using the same. In today's world machine learning has become an important part of everything we are a part of. It helps various enterprises for the development of new products and also helps to have an idea about the customer trends. Many MNC's have made ML an important part of their organisation.

The objective of a flight ticket price prediction system is to predict the price of flight tickets based on various factors, such as the airline, departure and destination cities, day of the week, time of booking, and other features like duration, layovers, and weather. Here is an overview of the steps involved in creating such a system:

1. Dataset Selection

To predict flight ticket prices, you need a dataset containing historical data on flights with features that affect the price, such as:

Flight Date & Time: Departure and arrival times.
Airline: The airline company.
Departure and Arrival Cities: Locations of the flight's start and end.
Class: Economy, Business, First Class.
Duration: Length of the flight.
Number of Stops: Direct or layovers.
Price: The price of the ticket (target variable).

Sources for Datasets:

Kaggle Flight Price Prediction Dataset: A widely used dataset containing data on flight prices and their associated features. It can be found on Kaggle or other open data platforms.

The learning process for an ML algorithm begins and ends with data. Therefore, data is the most important part of a model. This data has an effect on future predictions and decisions. The primary objective of ML is that a computer system learns automatically without human assistance.

Machine Learning is categorized mainly into 4 approaches based on how an algorithm goes through the process of learning.

Supervised machine learning: Analysis of labelled data and then training on the same to predict future events. This system can provide output for new inputs if sufficient training is done.

Unsupervised machine learning: The information used in training is not labelled and neither classified. This system is used to study different patterns and connections in data.

Semi-supervised machine learning: Falls in

between the above two as it uses small amount of labelled data and larger amount of the unlabelled data.

Reinforcement machine learning: In this system the program learns to behave in a particular environment by performing actions and then seeing the results.

Decision tree is a powerful tool used for classification and prediction. It resembles the shape of a flowchart. They classify instances from the root node towards the leaf node. A condition is specified at each node(internal node) based on which the tree further divides into edges.

II. LITERATURE SURVEY

Travelling by the mode of airplanes from one place to another is perhaps the easiest and most convenient ways but selecting the right type of flight with minimum ticket expense and the perfect route is a cumbersome process for a customer. Hence sophisticated Artificial Intelligence technique called Machine Learning provides a suitable and convenient platform to the users for selecting the right type of flight with minimum price and suitable route. Flight ticket price prediction is a challenging task as the various factors/variables determining the cost can change constantly over time.[1]

Flight prices can vary drastically depending on whether the day was a festival, weekday, weekend, peak season etc. Traditional factor like the total distance between source and destination is not the sole factor which determines the price but has a broader aspect depending on a number of factors. Generally the flights that have departure during weekends have a higher price than those which departure during weekdays [5].Also flight prices during festive season are generally lower. The route and number of stops also play a major role in determining the price of the flights. Features like total number of stops, Duration hours, journey day are some of the most important features for the prediction purpose as the price of the flight tickets is highly dependent on these types of features. To find the most important features correlation heatmap is used which gives the correlation values and higher the correlation value between two features higher are the chances of one of them being removed.

Various Machine Learning algorithms have been used to predict the flight ticket price whether on international level or national level. A number of datasets have also been used for the same although the main essence of prediction is the same for all which deals with proper preprocessing and feature selection. Various algorithms like Linear Regression, Decision tree, Support Vector Machines (SVMs) and Neural network are commonly used for building this model. Among all the algorithms mentioned above, Random forest gives the best accuracy. SVMs fail to give a desired accuracy on regression analysis however it can be used for classifying the price as high or low.[3]

The following algorithms were tested for accuracy among which Decision tress proved to be the best:

Linear Regression

Decision tree Regressor

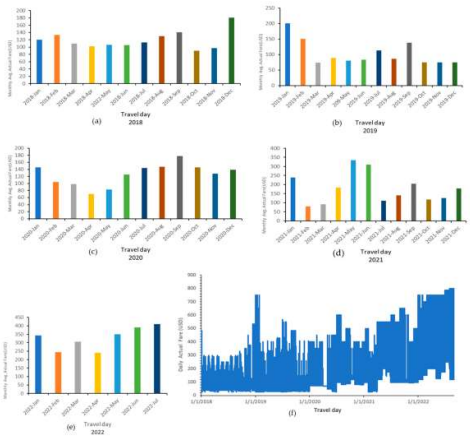
Random forest

Linear Regression however does not prove to be a suitable algorithm and only gives an accuracy of 61 percent. As the numbers of features are more in number and complex in nature, linear regression fails to give the expected accuracy. Decision tree however gives an accuracy of 72 percent which is quite good for a dataset of this size.[2]

Random forest is the most suitable algorithm as it uses multiple decision trees to get a final output and thus results in maximum accuracy among the above mentioned algorithms. It gives an accuracy of 79 percent and along with hyper-parameter tuning the accuracy jumps to almost 81 percent. Random forest has almost equivalent parameters as a decision tree.

In order to find the best features for fitting the model, feature selection is very important as it investigates the degree of impact of each feature on prediction output. Correlation between the features determines how closely the features are related to one another. One of the two highly correlated features are removed so as to decrease the size of features and decrease the complexity in fitting the model.

RESULTS



CONCLUSIONS

The analysis of world population trends using machine learning provides valuable insights into the patterns and factors influencing global population growth. Through the application of various machine learning techniques, such as regression models, time series analysis, and classification algorithms, we can predict population changes, identify key drivers of growth, and forecast future trends.

Key takeaways from this analysis include:

1. Data-Driven Insights: By utilizing large datasets on population, socio-economic factors, birth rates, death rates, migration patterns, and government policies, machine learning models can help uncover hidden patterns and relationships in the data. This can lead to more accurate predictions and better-informed decision-making for governments and organizations.
2. Predictive Power: Machine learning models like linear regression, decision trees, and more advanced models like neural networks can be used to predict future population trends based on historical data. These predictions are crucial for long-term planning in areas such as urban development, resource allocation, healthcare, and environmental sustainability.
3. Importance of Data Quality: The accuracy and performance of machine learning models heavily depend on the quality of the data used. Clean, comprehensive, and well-structured data is essential for making reliable predictions. Issues such as missing data, outliers, or incorrect data can undermine model performance and lead to inaccurate forecasts.
4. Challenges in Analysis: While machine learning can significantly enhance population analysis, challenges remain in ensuring data diversity and addressing biases that may arise due to socio-economic factors, geographic differences, or political influences. Additionally, certain external variables (e.g., pandemics or natural disasters) can drastically impact population growth, making long-term predictions inherently uncertain.

REFERENCES

1. T.Wang et al., "A Framework for Airfare Price Prediction: A Machine Learning Approach," 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 2019.

2. K. Tziridis, T. Kalampokas, G.A Papakostas and K.I Diamantaras, "Airfare prices prediction using machine learning techniques," 2017 25th European signal Processing Conference (EUSIPCO), 2017.

3. S Rajankar, N Sakharkar, "A Survey on Flight Pricing Prediction using Machine Learning", International Journal of Engineering, 2019.

4. Vinod Kimbhuane, Harshil Donga, Asutosh Trivedi, Sonam Mahajan and Viraj Mahajan, "Flight Fare Prediction System", May 2021

5. Prashant Kapri, Shubham Patane, Noopur Thanvi, Rashmi Thakur, "Predictive Model for Airlines Flight Delay and Pricing"