The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 127 - 131

Flight Delay Prediction Using Machine Learning

[1]K. Harshitha, [2]J. Lakshmi Priya, [3]Vasmitha, [4]Mounika
[1,2,3,4] UG Student, Department of Computer Science Engineering,
Dr K V Subba Reddy College Of Engineering For Women, Kurnool, Andhra Pradesh, India

Abstract
Predicting a delay class or value is the most common approach to the problem of predicting flight delays. Probabilistic delay predictions on an individual flight basis, on the other hand, provide insight into the uncertainty of delay predictions and can be of great benefit to the aviation industry. As a result, in order to predict flight delays at a European airport, this study employs Mixture Density Networks and Random Forest regression, two probabilistic forecasting algorithms. With a Mean Absolute Error of less than 15 minutes, the algorithms estimate the distribution of arrival and departure flight delays accurately. In a probabilistic flight-to-gate assignment problem, we incorporate these probabilistic predictions to demonstrate the value of the estimated delay distributions. Increasing the robustness of flight-to-gate assignments is the goal of this problem. When compared to a deterministic flight-to-gate assignment model, our proposed flight-to-gate assignment model reduces the number of conflicted aircraft by up to 74% when probabilistic delay predictions are taken into consideration. Overall, the findings demonstrate the value of probabilistic forecasting for stable airport operations optimization

1. Introduction

Due to the associated financial losses that the aviation industry is experiencing, flight delays have become a very important topic for air transportation all over the world. The United States' Bureau of Transportation Statistics (BTS) reports that more than 20% of flights in the country were delayed in 2018, which had a significant economic impact of 41 billion US dollars. Both the airlines and the passengers suffer as a result of these delays. The end result is a longer travel time, more money spent on food and lodging, and ultimately more stress for passengers. The airlines are the victims of additional costs related to their crews, aircraft repositioning, fuel consumption in an effort to shorten travel times, and many other factors, all of which affect their reputation and frequently lead to a decline in passenger demand. These delays can be caused by a variety of factors, including air congestion, weather, mechanical issues, difficulties boarding passengers, or simply the airlines' inability to meet the demand given their capacity. So, as a passenger, what can you do to avoid delayed flights? Is there a way to find out if your flight will be delayed before the departure boards show it? or prior to your boarding the plane? The response to these inquiries is possibly. You can attempt to predict a variety of flight delays by utilizing two algorithms. Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals. Of course, all of these different algorithms will have pitfalls and a certain degree of accuracy that will be associated with the data they are fed. learn from past mistakes. Without relying on a predetermined equation as a model, machine learning algorithms "learn"

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 127 - 131

information directly from data using computational methods. As more samples are made available for learning, the algorithms adjust their performance. A model is developed using supervised machine learning to make predictions in the face of uncertainty based on evidence.

A supervised learning algorithm trains a model to make reasonable predictions for the response to new data using a known set of input data and known responses to the data output. If you have known data for the output you want to estimate, use supervised learning. The development of a machine learning model for supervised learning makes use of classification and regression techniques.

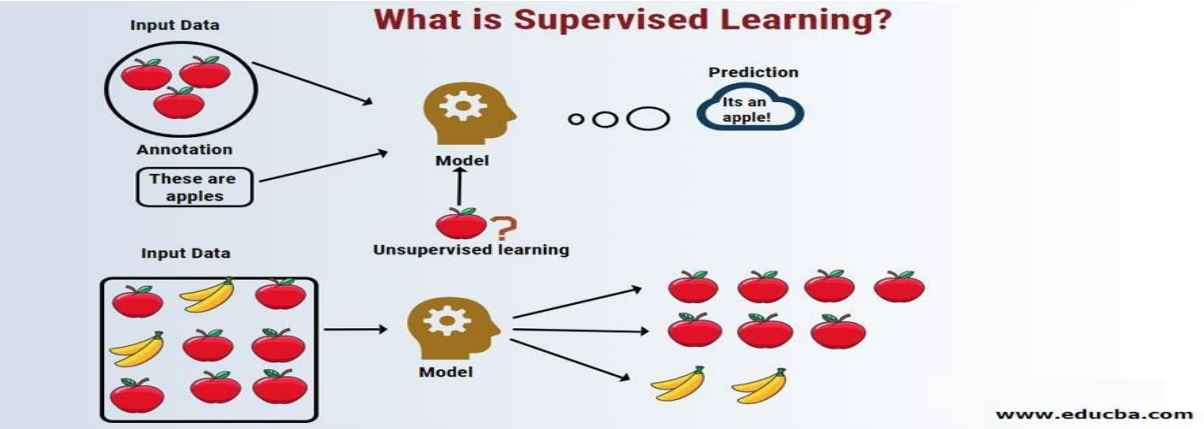

Fig.1 Supervised Machine Learning

2. Literature Review

supervised machine learning includes regression techniques. They aid in the interpretation or prediction of a specific numerical value based on previous data, such as the price of an asset based on previous pricing data for properties that are comparable.

The simplest approach is linear regression, in which the data set is modeled using the line's mathematical equation ($y = m * x + b$). By calculating the position and slope of a line that minimizes the total distance between all data points and the line, we train a linear regression model with multiple data pairs (x, y). To put it another way, we determine the y-intercept (B) and slope (M) of a line that is closest to the data's observations.

Let's take a look at a more concrete illustration of linear regression. By combining the age of the building, the number of stories, square feet, and the number of wall devices plugged in, I once used linear regression to predict the energy consumption of some buildings. In another category of supervised machine learning, classification methods predict or explain a class value. For instance, they can assist in determining whether an online customer will make a purchase. Yes or no can be the output: buyer or none at all. However, the classification methods are not restricted to just two classes. A classification technique, for instance, can assist in determining whether an image contains a car or a truck. Logistic regression, which looks like a regression method but is not, is the simplest classification algorithm. Based on one or more inputs, logistic regression calculates the probability that an event will occur. k-means and k-medoids, hierarchical clustering, Gaussian mixture models, hidden Markov

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 127 - 131

models, self-organizing maps, fuzzy C-means clustering, and subtractive clustering are all common algorithms for clustering.

Fig.2 Clustering

3. Proposed System

Python is a popular high-level, general-purpose programming language. Guido van Rossum developed it in 1991, and the Python Software Foundation added to it later. Programmers can express their ideas in fewer lines of code thanks to its syntax, which was made with code readability in mind when it was designed.

The programming language Python allows you to work more quickly and efficiently integrate systems. There are two main versions of Python: Python versions 2 and 3 They are very different. HTML, or Hypertext Markup Language, is one of the world wide web's languages. It lets users create web pages with text, graphics, and hyperlinks to other websites.

HTML is an ISO Standard application, not a programming language. 8879, an adaptation of SGML (Standard Generalized Markup Language) for the Web and hypertext. Hypertext's concept of linking points together. We can sort through the information according to our preferences and interests. A markup language is merely a collection of elements that should be displayed. Hyperlinks are works that are highlighted and link to other documents or portions of the same document

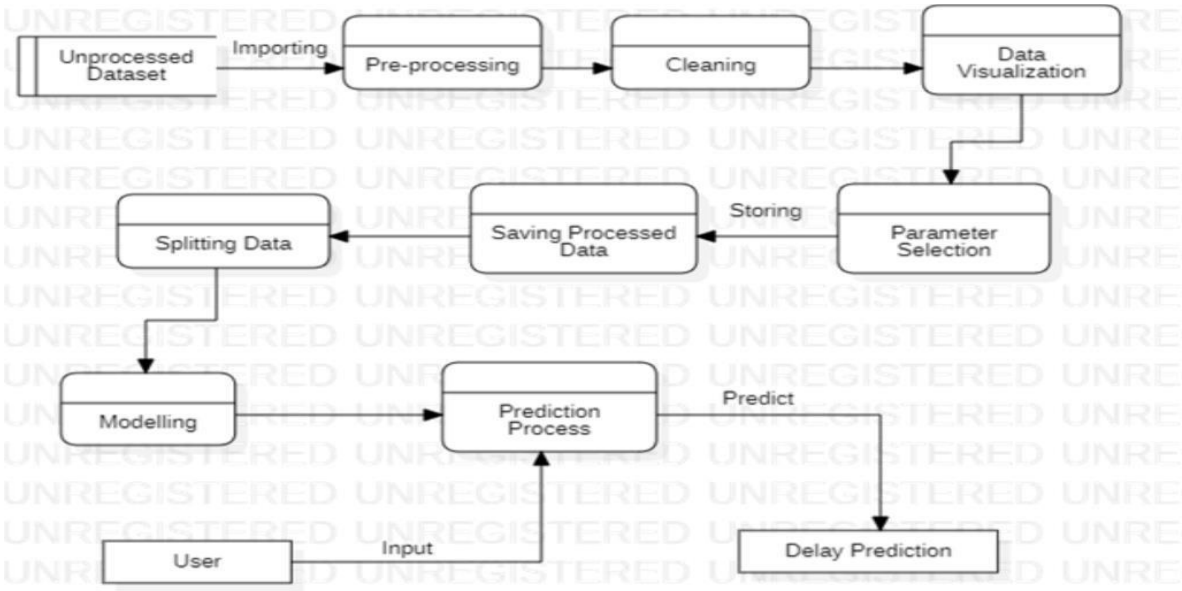

Fig.3 Proposed Architecture

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 127 - 131

We can collect the unprocessed data and importing the pre processing technique and clean the data and create the some graphs to understand the model by taking different columns and convert the character data into numerical data if any present so that the prediction became easy and then we can split the data into two parts that is test data and train data the test data should always be less then the train data so that we are taking the test data is 20% and the train data is 80% after that we can model the data by using the some classification algorithms like Decision tree classifier and also use logistic algorithm , support vector etc. Then can find the accuracy. With the help of the flask we can create the web page in that the user can give the input and the prediction is processed and give the output that flight will come on time or not

Fig.4 Output Of The Project

5.Conclusion

We use demand data, weather data, and flight data to predict flight delay in this project. According to our findings, the Random Forest approach outperforms the SVM model in terms of performance. The SVM model is, for some reason, very time-consuming and does not always result in better outcomes. In the end, 91% of non-delayed flights are correctly predicted by our model. However, only 41% of the time, the delayed flights are correctly predicted. Consequently, our existing data sources may not yet have revealed additional features related to the causes of flight delays. We can see in the second part of the project that the volume of concurrently published tweets, as well as the sentiment and objectivity of those tweets, can be used to predict flight delay patterns. It's not unreasonable for this; On Twitter, individuals frequently discuss delays at airports; As the delays worsen, it stands to reason that these posts would become more frequent and more profoundly emotional. We are unable to construct a robust model or determine the impact of related factors and chance on these results without additional data. However, these outcomes might work as a proof of concept. It may be possible to utilize tweets on a regular basis to gain an understanding of concurrent airline delays and traffic patterns, which could be useful in a number of situations.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 127 - 131

References

1. A. B. Guy, "Flight delays cost $32.9 billion, passengers foot half the bill". [Online] Available: https://news.berkeley.edu/2010/10/18/flight_delays/3/. [Accessed on June 2017].
2. M. Abdel-Aty, C. Lee, Y. Bai, X. Li and M. Michalak, "Detecting periodic patterns of arrival delay", Journal of Air Transport Management,, Volume 13(6), pp. 355– 361, November, 2007.
3. S. AhmadBeygi, A. Cohn and M. Lapp, "Decreasing Airline Delay Propagation By Re-Allocating Scheduled Slack", Annual Conference, Boston, 2008.
4. A. A. Simmons, "Flight Delay Forecast due to Weather Using Data Mining",
5. M.S. Disseration, University of the Basque Country, Department of Computer Science, 2015.
6. S. Choi, Y. J. Kim, S. Briceno and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms", Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th, Sacramento, CA, USA, 2016.
7. L. Schaefer and D. Millner, "Flight Delay Propagation Analysis with the Detailed Policy Assessment Tool", Man and Cybernetics Conference, Tucson, AZ, 2001.
8. B. Liu "Sentiment Analysis and Opinion Mining Synthesis", Morgan & Claypool Publishers, p. 167, 2012. 55
9. Statistical Computing Statistical Graphics. [Online]. Available: http://statcomputing.org/dataexpo/2009/the-data.html. [Accessed on April 2017].
10. FAA Operations & Performance Data. [Online]. Available: https://aspm.faa.gov/.[Accessed on April 2017].