Team 12

Akshay Reddy
Paleti Krishnasai
Sai Gowtham Babu
Rituraj Choudhury

Application Of Machine Learning Algorithms To Predict Flight Delays

Growth in aviation industry has resulted in air-traffic congestion causing flight delays. Flight delays not only have economic impact but also harmful environmental effects. Air-traffic management is becoming increasingly challenging. In this project we apply machine learning algorithms like decision tree and logistic regression to predict if a given flight's arrival will be delayed or not.

Introduction

Over the past few years, air travel has been increasingly preferred among travellers, mainly because of its speed, safety and in some cases comfort. This has led to increased growth in the air traffic . Increase in air traffic has also resulted in massive levels of aircraft delays on the ground and in the air. These delays are responsible for large economic losses to MNCs.In this fast paced world , everyone wishes to get things done quickly , delays in airports will cause airlines to loose business. It will be an advantage to airline insdustry if flight delays are predicted well in advance.

Machine learning is a promising field that explores algorithm development which learn from existing data and provide predictions based on it. In this project, Decision Tree

Algorithm and Logistic Regression have been applied on an airline dataset to predict flight delays.

The input to the algorithms are rows of feature vector comprising of features like departure date, departure delay, distance between the two airports, scheduled arrival time etc. . A flight is considered to be delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, a comparitive study is performed on the results obtained using decision tree classifier and logistic regression.Finally, the project is integrated to web based application.

Word-Cloud of the project.

Literature Survey

There are several works in the literature that focus on air-traffic management and optimization. Studies on flight systems are increasing using machine learning methods. The commonly used methods include Neural networks, k-Nearest Neighbor, SVM, fuzzy

logic etc. These machine learning methods are mainly used for classification and prediction. A few of the works are presented below.

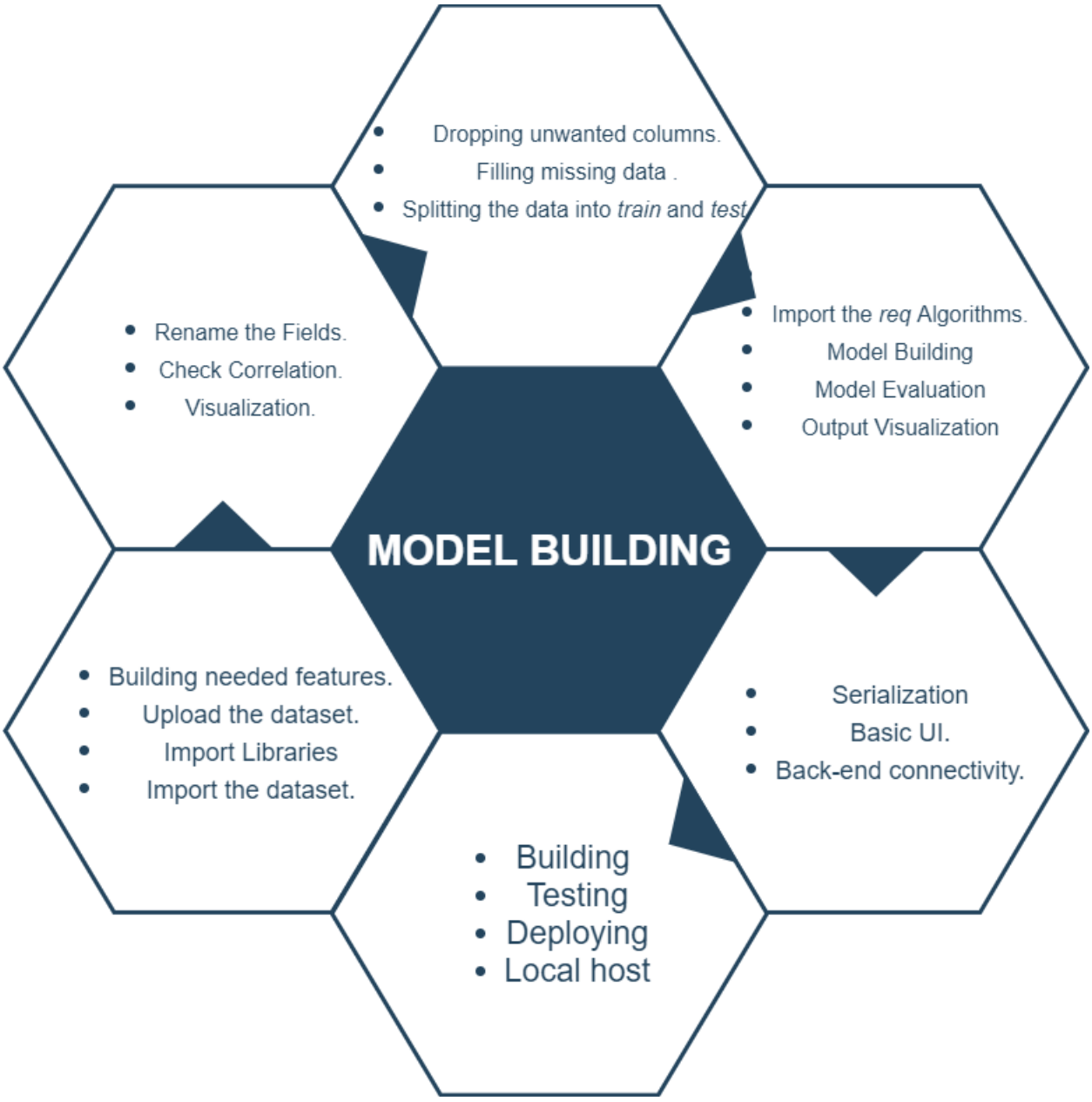
Root delay has been predicted using random forest by Rebollo et al.(1). The authors compared their approach with regression models to predict root delay in airports of the United States considering time horizons of 2, 4, 6 and 24 hours. There was an increase in the test errors as the forecast horizon increased.

An adaptive network based on fuzzy inference system was created by Khanmohammadi et al.(2) to predict root delay. The predictions were used as an input for a fuzzy decision-making method to structure arrivals at JFK International Airport in New York.

Reinforcement learning algorithm has been used by Balakrishna et al.(3,4) to predict taxi-out delays. Markov decision process has been used to model the problem and it was implemented by a machine learning algorithm. When running the model 15 minutes before the scheduled time of departure, authors achieved good performances at JFK International Airport in New York and Tampa Bay International Airport.

A recommendation system to predict delay at select airports based on propagation effects has been developed by Lu et al.(5) . The prediction was based on the k-Nearest Neighbor algorithm and used historical data to recognize similar situations in the past. The authors observed fast response time and easy, logical comprehension as the main advantages of their method.

Theoritical Design



Experimental Investigations

Upon applying Logistic Regression and Decision Tree Classifier Algorithms, it is observed that the accuracy of Decision Tree Classifier is higher when compared with Logistic Regression.

Results

```
In [31]: model.score(x_test, y_test)
Out[31]: 0.965

         ACCURACY SCORE

In [32]: accuracy_score(y_test, y_pred)
Out[32]: 0.965

         F1 SCORE

In [33]: score = f1_score(y_test, y_pred, average='micro')
         print('F-Measure: %.3f' % score)
         F-Measure: 0.965

         PRECISION

In [34]: precision = precision_score(y_test, y_pred, labels=[1,2], average='micro')
         print('Precision: %.3f' % precision)
         Precision: 0.833

         RECALL OR SENSITIVITY

In [35]: recall = recall_score(y_test, y_pred, labels=[1,2], average='micro')
         print('Recall: %.3f' % recall)
         Recall: 0.870

         CONFUSION MATRIX

In [36]: results = confusion_matrix(y_test, y_pred)
         print(results)
         [[173   4]
          [  3  20]]

         CROSS_VAL_SCORE

In [37]: cross_val_score(model,x,y.ravel(),cv=5)
Out[37]: array([0.75      , 0.93      , 0.905     , 0.93      , 0.89447236])

In [38]: cross_val_score(model,x,y.ravel(),cv=10).mean()
Out[38]: 0.9409292929292927
```

Advantages

- Easily accessible.
- Instantaenious and accurate results.
- Simple to use UI.

Disadvantages

- The data for certain fields will only be known with close proximity to

departure time.

Applications

In general Decision Trees used in various fields such as stock trading , medtech(eg:Diagnosing Xrays of Thoriac Pathology)etc. In this Project Decision Tree is used to Predict Flight Arrival Delays.

This Project can be used in several airline agencies and the expected delays in the flights can be communicated to the passengers .

Conclusion

In this project , machine learning algorithms namely Logistic Regression and Decision Tree have been used to predict Flight Arrival Delays .
Airline Dataset comprising of various attributes such as scheduled arrival time, distance between airports etc ,.is considered for experimentation.
The dataset is preprocessed before applying the model so as to improve the performance of the model.
The model building steps were implemented in python on jupyter notebook.
It is observed that Decision Tree gives a better accuracy when compared to Logistic Regression.

Future Scope

In order to further improve the performance of the model, several other factors like weather conditions , air traffic details etc., can also be considered in the dataset.

References

[1] J. J. Rebollo and H. Balakrishnan. Characterization and prediction of air traffic delays. Transportation Research Part C: Emerging Technologies, 44(Supplement C):231–241, July 2014.

[2] S. Khanmohammadi, C. A. Chou, H. W. Lewis, and D. Elias. A systems approach for scheduling aircraft landings in JFK airport. In 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pages 1578–1585, July 2014.

[3] P. Balakrishna, R. Ganesan, and L. Sherry. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. Transportation Research Part C: Emerging Technologies, 18(6):950–962, Dec. 2010.

[4] P. Balakrishna, R. Ganesan, L. Sherry, and B. S. Levy. Estimating Taxi-out times with a reinforcement learning algorithm. In 2008 IEEE/AIAA 27th Digital Avionics Systems Conference, pages 3.D.3–1–3.D.3–12, Oct. 2008.

[5] L. Zonglei, W. Jiandong, and Z. Guansheng. A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning. In 2008 International Symposium on Knowledge Acquisition and Modeling, pages 589–592, Dec. 2008.

Appendix

Project
(contains notebooks, dataset and Flask Folder)