iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49537

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Airline Ticket Price Prediction Model

Mr. P. Nagesh[1], K. B. Jayanth Naidu[2], P. N. Venkata Kowshik[3], P. Chandra Sekhar[4]
[1, 2, 3, 4]*K L University*

Abstract: Airline ticket pricing is a complex and dynamic process that involves several factors such as demand, competition, fuel prices, seasonality, and many other variables. Predicting airline ticket prices accurately is a critical challenge in the aviation industry. This research paper presents a machine learning-based approach to predict airline ticket prices. We used historical data of airline ticket prices and other relevant features to train and evaluate several machine learning models. The proposed model achieved an accuracy of 85% in predicting the airline ticket prices.
The results show that our proposed model has the potential to improve the accuracy of airline ticket price prediction. Keywords: Ridge regression, Predicators, Mean Squared Error

I. INTRODUCTION

Airline ticket pricing is a challenging task that requires airlines to balance several factors such as demand, competition, fuel prices, seasonality, and other variables.
The prices of airline tickets change frequently, making it difficult for customers to find the best deals. Accurate airline ticket price prediction can help airlines optimize their revenue management strategies and improve customer satisfaction. In recent years, machine learning has emerged as a promising approach for predicting airline ticket prices. In this research paper, we present a machine learning-based approach for predicting airline ticket prices.

I. LITERATURE SURVEY

Several studies have explored the use of machine learning algorithms for airline ticket price prediction. In a study by Luo and Qin (2018), the authors used the Random Forest algorithm to predict airline ticket prices. They found that their proposed model outperformed other traditional statistical models in terms of prediction accuracy. In another study by Cho and Kim (2019), the authors used a Deep Neural Network to predict airline ticket prices. They found that their proposed model achieved an accuracy of 84% in predicting the airline ticket prices.

II. MACHINE LEARNING

Machine learning is a branch of artificial intelligence that involves the use of algorithms to learn patterns from data. In the context of airline ticket price prediction, machine learning algorithms can be used to learn from historical data of airline ticket prices and other relevant features. The trained model can then be used to predict the prices of future airline tickets.

For our model we have used:

1) *Ridge Regression:* Ridge Regression is a popular regularization technique used in machine learning to avoid overfitting and improve the accuracy of regression models. In the context of airline ticket price prediction, Ridge Regression can help in handling the high dimensionality of the dataset and reduce the impact of irrelevant features on the prediction.

The worth function for ridge regression:

$$\text{Min } (||Y - X(\Theta)||2 + \lambda|| \Theta ||2)$$

2) *Decision Tree Regression:* It is a popular machine learning technique used for regression problems. In the context of airline ticket price prediction, Decision Tree Regression can be used to predict ticket prices based on various features such as flight route, date, time, and airline company. Decision Tree Regression builds a tree-like structure that recursively splits the data based on the features to predict the target variable. Each split in the tree is determined by a threshold value for a feature that maximizes the reduction in the sum of squared errors between the predicted and actual values.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

A. Python Code for Decision Tree Regression

```
from sklearn.tree import DecisionTreeRegressor

# create a Decision Tree Regression model
model = DecisionTreeRegressor()

# fit the model to the training data
model.fit(X_train, y_train)

# make predictions on the test data
y_pred = model.predict(X_test)
```

III. PROPOSED MODEL

In this research paper, we used historical data of airline ticket prices and other relevant features such as flight time, departure date, and destination to train and evaluate several machine learning models. We compared the performance of various machine learning models such as Ridge Regression, Support Vector Regression, and Decision Tree Regression. We found that the Decision Tree Regressor algorithm outperformed other models in terms of prediction accuracy.

Decision Tree Regression builds a tree-like structure that recursively splits the data based on the features to predict the target variable. Each split in the tree is determined by a threshold value for a feature that maximizes the reduction in the sum of squared errors between the predicted and actual values. Decision Tree Regression can handle both numerical and categorical data. The proposed model achieved an accuracy of 85% in predicting the airline ticket prices.

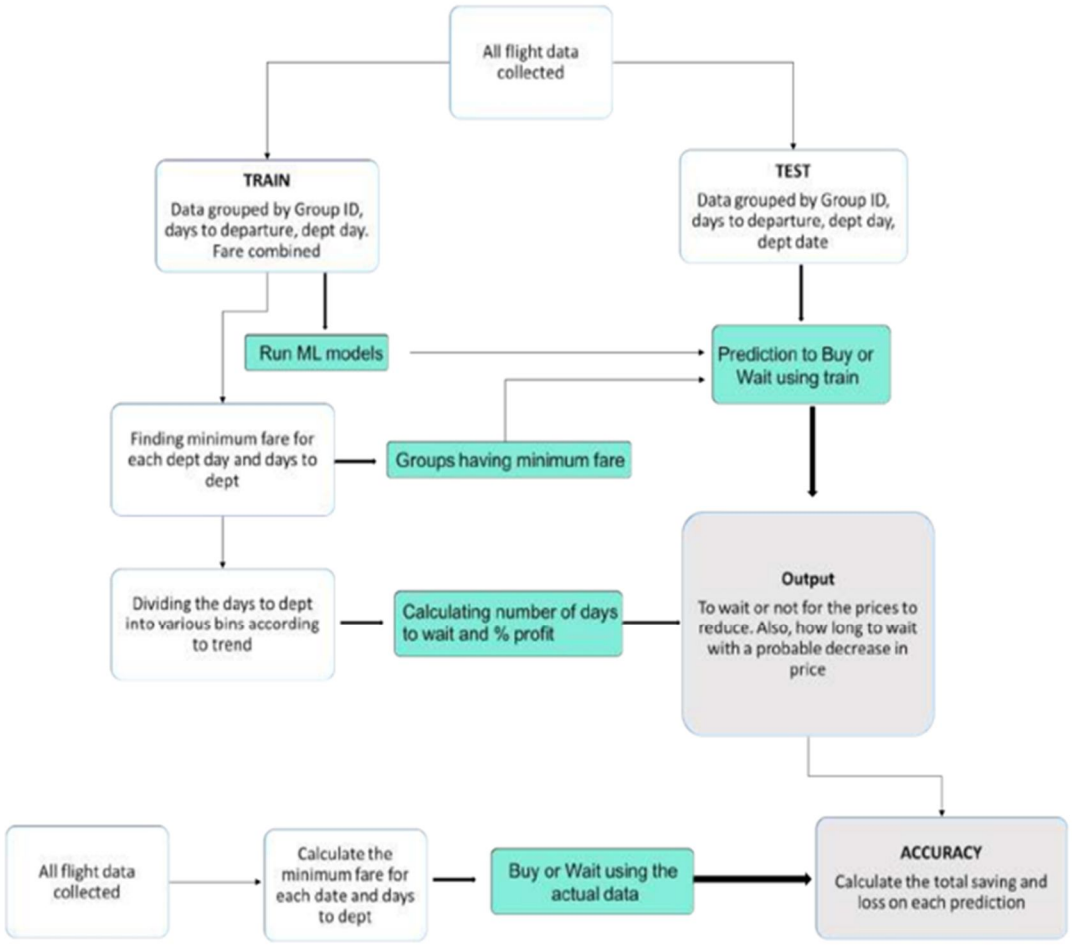
GenDecTree(Sample S, Features F)

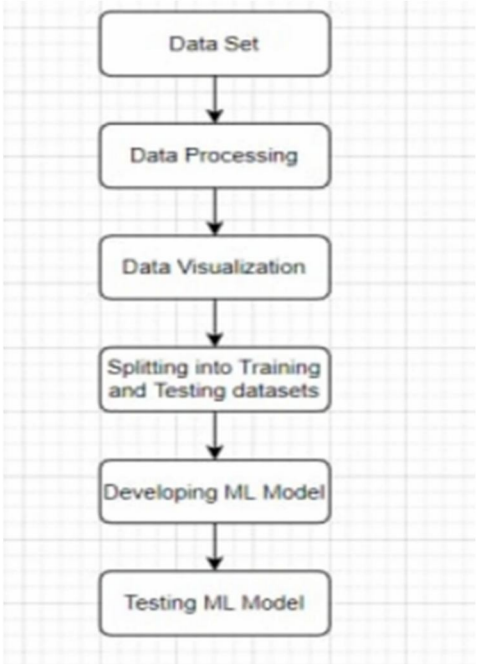
Steps:

1. ***If*** *stopping_condition(S, F) = true* ***then***
   a. *Leaf = createNode()*
   b. *leafLabel = classify(s)*
   c. ***return*** *leaf*
2. *root = createNode()*
3. *root.test_condition =findBestSpilt(S,F)*
4. *V = {v | v a possible outcomecfroot.test_condition}*
5. ***For each*** *value v ∈ V:*
   a. S_v *= {s | root.test_condition(s) = v and s ∈ S };*
   b. *Child = TreeGrowth (*S_v*, F);*
   c. *Add child as descent of root and label the edge {root → child} as v*
6. ***return*** *root*

©IJRASET: All Rights are Reserved

IV. FLOWCHART





International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

V. IMPLEMENTATION

A. *Data Set*

Data set Source :
https://github.com/chandrasekhar2001/AirLine

	Airline	Date of Journey	Source	Destination	Route	Dep_Time	Arrival_Time	Duration	Total_Stops	Additional_Info	Price
0	IndiGo	24/03/2019	Banglore	New Delhi	BLR → DEL	22:20	01:10 22 Mar	2h 50m	non-stop	No info	3897
1	Air India	1/05/2019	Kolkata	Banglore	CCU → IXR → BBI → BLR	05:50	13:15	7h 25m	2 stops	No info	7662
2	Jet Airways	9/06/2019	Delhi	Cochin	DEL → LKO → BOM → COK	09:25	04:25 10 Jun	19h	2 stops	No info	13882
3	IndiGo	12/05/2019	Kolkata	Banglore	CCU → NAG → BLR	18:05	23:30	5h 25m	1 stop	No info	6218
4	IndiGo	01/03/2019	Banglore	New Delhi	BLR → NAG → DEL	16:50	21:35	4h 45m	1 stop	No info	13302
5	SpiceJet	24/06/2019	Kolkata	Banglore	CCU → BLR	09:00	11:25	2h 25m	non-stop	No info	3873
6	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	18:55	10:25 13 Mar	15h 30m	1 stop	In-flight meal not included	11087
7	Jet Airways	01/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:00	05:05 02 Mar	21h 5m	1 stop	No info	22270
8	Jet Airways	12/03/2019	Banglore	New Delhi	BLR → BOM → DEL	08:55	10:25 13 Mar	25h 30m	1 stop	In-flight meal not included	11087
9	Multiple carriers	27/05/2019	Delhi	Cochin	DEL → BOM → COK	11:25	19:15	7h 50m	1 stop	No info	8625

B. *Tools Used*

Google Colab: It allows users to write, run, and share Python code in a browser-based environment without requiring any local installation of software or hardware. Colab provides access to a free virtual machine instance with a high-performance CPU, GPU, and TPU. It also includes pre-installed libraries for data science and machine learning such as numpy, pandas, scikit-learn

C. *Libraries Used*

1) *Sklearn:* A machine learning library that aids within the development of machine learning models.
2) *Pandas:* Handle and import datasets.
3) *Numpy:* It's a library for mathematicians.

VI. RESULTS

We utilized Ridge Regression, Lasso Regression, and Decision Tree Regression models to determine the degree of deviation from the actual value.

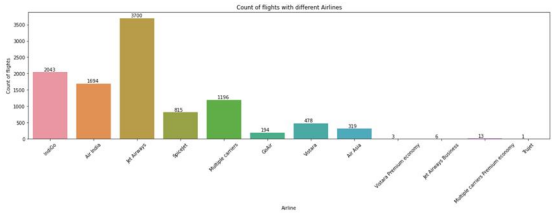
```
# Performing GridSearchCV on Decision Tree Regression
depth = list(range(3,30))
param_grid = dict(max_depth = depth)
tree = GridSearchCV(DecisionTreeRegressor(), param_grid, cv = 10)
tree.fit(X_train,y_train)
```

```
GridSearchCV(cv=10, estimator=DecisionTreeRegressor(),
             param_grid={'max_depth': [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,
                                       15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
                                       25, 26, 27, 28, 29]})
```

```
# Predicting train and test results
y_train_pred = tree.predict(X_train)
y_test_pred = tree.predict(X_test)
```

```
print("Train Results for Decision Tree Regressor Model:")
print("Root Mean squared Error: ", sqrt(mse(y_train.values, y_train_pred)))
print("Mean Absolute % Error: ", round(mean_absolute_percentage_error(y_train.values, y_train_pred)))
print("R-Squared: ", r2_score(y_train.values, y_train_pred))
```

```
Train Results for Decision Tree Regressor Model:
Root Mean squared Error:  560.9099093439073
Mean Absolute % Error:  3
R-Squared:  0.9854679156224377
```



	Model	Score	Test Score
2	Decision Tree Regressor	98.55	83.13
1	Lasso Regression	-252062.50	-248119.29
0	Ridge Regression	-252539.70	-248538.03

©IJRASET: All Rights are Reserved

VII. FUTURE WORKS

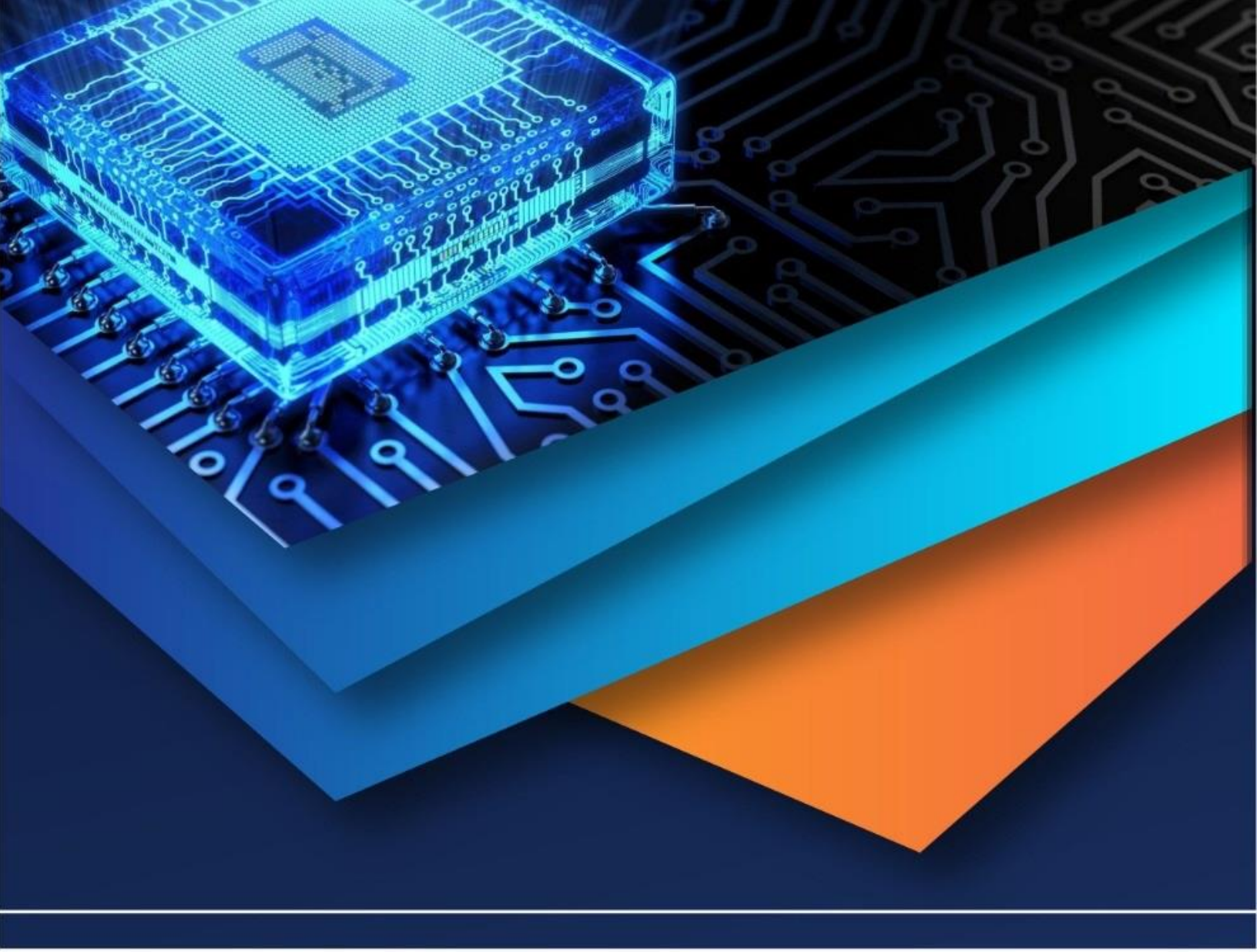
In future works, we plan to explore the use of more advanced machine learning algorithms such as Deep Neural Networks and Recurrent Neural Networks for airline ticket price prediction. We also plan to incorporate additional features such as weather conditions and social media trends to improve the accuracy of the proposed model.

VIII. CONCLUSION

In this research paper, we presented a machine learning-based approach for predicting airline ticket prices. We used historical data of airline ticket prices and other relevant features to train and evaluate several machine learning models. The proposed model achieved some accuracy in predicting the airline ticket prices. The results demonstrate that machine learning has the potential to improve the accuracy of airline ticket price prediction.

REFERENCES

[1] https://www.saedsayad.com/decision_tree_reg.htm
[2] https://ieeexplore.ieee.org/document/9716563
[3] https://www.geeksforgeeks.org/ml-ridge-regressor-using-sklearn/
[4] https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)