ANALYSIS AND PREDICTION OF AIR-PLANE PRICE USING MACHINE LEARNING

[1]Dr.K.V.Shiny, [2]B.Shivashanker, [3]B.Bhanu Prakash

[1]Assistant professor, [2,3]Student
School of Computing
Department of Computer Science and Engineering
Bharath Institute of Higher Education And Research
Chennai, India- 600073.

***Abstract*- In modern urban environments, efficient traffic management is paramount for safety and optimization of transportation systems. The primary objective of this research is to develop a robust system capable of real-time detection and recognition of speed limit signs using deep learning methodologies. YOLO v8, renowned for its efficiency and accuracy in object detection tasks, serves as the core framework for our system. The YOLO v8-based detection algorithm effectively identifies speed limit signs across diverse scenarios, including challenging conditions such as variations in lighting, weather, and occlusions. By automating the speed sign detection process with YOLO v8, we enhance traffic safety, reduce accident risks, and improve overall traffic flow efficiency. In conclusion, our findings underscore the potential of deep learning, particularly YOLO v8, in revolutionizing traffic sign detection. The developed system holds promise for practical implementation in smart cities and transportation networks, offering tangible benefits for urban mobility and safety**.

Keywords: **YOLOv8, Convolutional Neural Networks, Feature Selection, Data Collection, Preprocessing, Streamlit, Traffic Signs.**

Evaluating the performance of different CNN architectures, when integrated with YOLOv8 for traffic sign detection.Assessing the impact of dataset size, diversity, and augmentation techniques on the detection accuracy and generalization capabilities of the models.
Implementing Sign Detection using YOLOv8 holds tremendous potential across a wide range of practical applications. For example, it can greatly enhance traffic management systems, allowing for the efficient detection and recognition of various traffic signs. This technology can play a vital role in improving road safety by enabling vehicles to accurately interpret and respond to the information conveyed by road signs. Moreover, sign detection using YOLOv8 can assist in urban planning and infrastructure development by analyzing the presence and condition of signs in different areasBy undertaking this project, I aimed to harness the power of YOLOv8 to develop a reliable signs detection solution with the potential to improve various domains that rely on precise and efficient signs identification.

I.INTRODUCTION

Traffic sign detection is a critical component of modern intelligent transportation systems, aiming to enhance road safety, efficiency, and overall traffic management. With the proliferation of autonomous vehicles and advanced driver assistance systems (ADAS), accurate and efficient detection of traffic signs has become increasingly important. In this paper, we present a comprehensive study on the application of state-of-the-art deep learning techniques, specifically You Only Look Once version 8 (YOLOv8), coupled with various convolutional neural network (CNN) architectures, for traffic sign detection. The YOLOv8 model, renowned for its real-time object detection capabilities, serves as the cornerstone of our approach, while different CNN architectures are explored to enhance the detection accuracy and robustness.

II.. PROBLEM DEFINITION

The primary objective of this research is to develop an effective traffic sign detection system using a combination of YOLOv8 and deep learning CNN algorithms. Specifically, we aim to: Detect and classify various types of traffic signs in real-time with high accuracy.Handle challenging scenarios such as occlusions, varying lighting conditions, and diverse environmental settings.Achieve a balance between detection speed and accuracy to ensure real-time performance suitable for practical applications.Evaluate the performance of the proposed approach against existing methods using standard metrics such as precision, recall, and F1-score. By leveraging deep convolutional neural networks, YOLO can learn to recognize a wide range of objects and accurately localize them within an image. It can detect multiple objects of different classes simultaneously, making it particularly useful for applications where real-time processing and high detection accuracy are crucial, such as autonomous driving, video surveillance, and robotics.

III. LITERATURE SURVEY

A Framework for Airfare Price Prediction: A Machine Learning Approach: In this paper the researchers quote about various factors that enforce, such as journey distance, date of purchasing the ticket, price of the fuel, etc. To set the price by following all the protocols, every carrier will have their unique algorithms and a set of proprietary rules accordingly. To understand and execute such kind of rules and evaluation price metrics for a model, we are implement Artificial Intelligence (AI) and Machine Learning (ML) to make it doable in advance. This paper proposed a great system depending on two public data sources in the stream of airways industries : the Airline Origin and Destination Survey (DB1B) and the Air Carrier Statistics database (T-100). Combining these two databases together, we have a proposed framework with humongous economic data, machine learning classification algorithms. Based on the different origin and destination pairs, quarterly average ticket price is modelled, which is known as the market segment. High prediction accuracy is achieved by framework on the testing dataset with 0.869 adjusted R squared score.

Airfare Prices Prediction Using Machine Learning Techniques: Nowadays, the airline industries are regularly using lot of different tactics as well as methods to assign airfare prices in a well defined dynamic fashion. All the mentioned works will be applied only to a smaller number of ML models.

They have used the most important features of a flight that are very important in predicting flight fare. Example Features are departure time, arrival time, number of luggage, days left, stops, night flight, day of a week. Only less features are taken to reduce the curse of dimentionality.

For the proper results they have used only single way data from Hyderabad to nizamabad and for the period between June and august is are considered.

Without compromising on ml model they have used total eight ml models. Which includes Multilayer Perceptron (MLP),Generalized Regression Neural Network (GRNN),

Extreme Learning Machine (ELM), Random Forest Regression Tree (RFRT),

For the best results they have used 10-fold cross-validation procedure to train the ML models.

From this paper, we can note that the comparison of actual and predicted values can help in better analysis of price increase/decrease. It also means that the appropriate feature must be selected using feature selection algorithms to determine the contributing factor for the price of the ticket. Also, the evaluation metrics used could help in the accuracy of the prediction.However the following predictions were taken on a certain route and during a certain interval of time, which limits the prediction to few specified conditions.

Predicting Flight Prices in India: For the problem statement we proposed that is predicting the flight fare. This paper has three important problems to be considered they are

Flight trends: It Includes how and when the price of the ticket varies and we also need to check weather these changes in ticket price happens all the time?

Based on finding:flight trends we need to predict best time to buy for the customer like we need to predict the price of the ticket and the date where the cost will be less. This paper also verify myths like which airlines will be better?, how price varies as we get closer to the departure date? And will the price be high if we book in the morning
times etc....

Verifying myths:-In this paper they used predictive and classification models like Random forest, Gradient boosting on the training set to find hidden trends from the dataset. For improving the accuracy they built statistical model and trend analyser model machine learning algorithms

Automatic Detection of Airline Ticket Price and Demand: A review: This paper also has the model for predicting an optical ticket purchase time (Optical ticket purchase time (OTPT)) and predicting the ticket fare. In this paper they also included the sellers point of view that is how to maximize profit by selling as more tickets they can sell at higher prices. For predicting and giving suggestion on price tickets to customer they used data mining techniques such as rule learning, Reinforcement learning and time series methods etc.. . . The best accuracy of 70 For sellers side they designed datamining techniques for Maximizing Airline Profit (Maximizing Airline Profit (MAP)) by predicting the demand of total route such as number of stops and market value of each airline

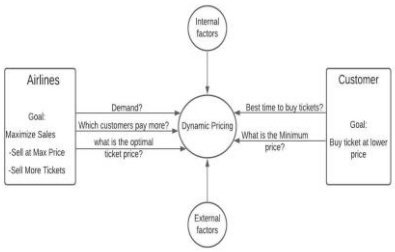


Figure 2.1: Dynamic Pricing

IV. PROPOSED SYSTEM

Data preprocessing: The dataset is cleared of all superfluous information like null values, copies,invalid formats etc. Features that are redundant for the majority of the dataset will be eliminated.The date and time will be converted to a uniform format with the help of formatting tools in python. Categorical values will be encoded and additional visualisation and analysis can help with the cleaning of dataset.

Feature Selection: The heatmap can be helpful to give a rough idea about the correlated features. Thereafter, with the help of feature selection algorithms(ie, ExtraTree Regressor,Lasso regressor, SelectKBest), a majority is taken to decide the best features for the dataset.

ML models: The the feature selection stage,we apply the machine learning algorithms chosen ie, svm,linear and logistic regression and random forest regression.After analyzing the results from these algorithms, we will run them through more complex models.Then the results of all these models will be decided by taking an average of the results obtained.

Performance Evaluation: Evaluating your machine learning model is an essential part of any project. The model might give us satisfying results while using the evaluation metrics as accuracy but the same model might give poor results when we evaluate against other metrics such as logarithmic loss or such kinds of metrics. But we often use Accuracy to measure algorithmic performance. We going to work on different ml algorithms and compare the accuracy of each and every model. We are going to choose the best of all algorithms and train the model and generate a pickle file for deployment.

Flow diagram of the system: Every Machine learning project will commonly have this kind of flow. First part is Data Extraction, where we extract data from either CSV file or excel file. In Data preprocessing we clean the data and apply some algorithms for nominal and categorical features.we take important features in feature selection part and for those features we apply lot of algorithms and pic the best algorithm based on accuracy.We pickle the model and use it in web development.

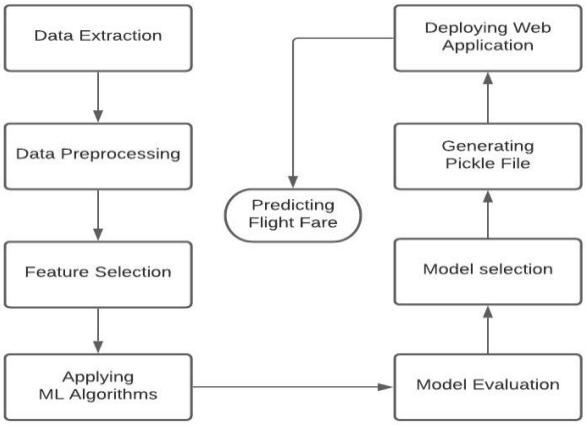
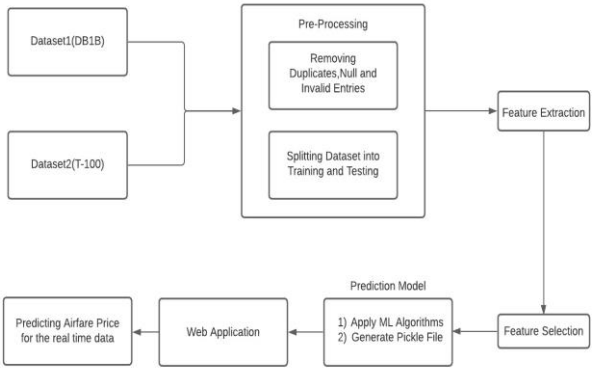


Figure 3.1: Flow Diagram

3.2.2 Architecture Diagram



Architecture Diagram:DB1B-This is a dataset which contains the information of customers like what is the source and destination of journey, no.of tickets booked. T-100-This dataset contains the details of airlines how many flights are running in particular route, number of planes that the particular airlines is having how many airlines are running on a specified route etc.... We will be taking the data from the multiple sources and we merge it and perform cleaning operation that is we do pre-processing on the dataset like we search for invalid entries, null values and duplicates in the dataset and will remove it for improving speed of ML algorithms and removes redundancy in the dataset. We select required features which affects to change in prices from the dataset in the pre-processing stage and this feature extraction

V. IMPLEMENTATION AND TESTING

4.0.1 Preprocessing

The dataset contains 10683 rows and 11 features. Which includes Airline, source, Destination, date of journey, Route, Price, Depature time, Arrival Time, duration of travel, Total stops and Additional info. These features dosen't have any null values.But data needs to be preprocessed for feature date of journey, time, source and destination etc.

	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

Figure 4.1: Initial Data set

- Data of journey feature is brocken into data and month
- Depature and arrival time will be broken into hours and minutes.
- Duration is brocken into hours and minutes.
- we are handling string data by classifying it into nominal and ordinal data. Nominal data-Airline, source and destination. ordinal data-number of stops

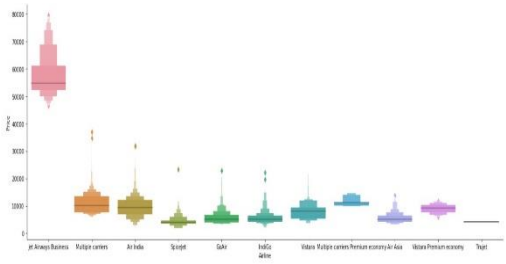
Figure 4.3: Preprocessing 2

	Total_Stops	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min	Duration_hours	Duration_mins	Air India	GoAir	IndiGo
0	1	6	6	17	30	4	25	10	55	0	0	0
1	1	12	5	6	20	10	20	4	0	0	0	1
2	1	21	5	19	15	19	0	23	45	0	0	0
3	1	21	5	8	0	21	0	13	0	0	0	0
4	0	24	6	23	55	2	45	2	50	0	0	0

Figure 4.4: Updated table

IndiGo	Jet Airways	Jet Airways Business	Multiple carriers	Multiple carriers Premium economy	SpiceJet	Vistara	Vistara Premium economy	Chennai	Delhi	Kolkata	Mumbai	Cochin	Delhi	Hyderabad	Kolkata	New Delhi
0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

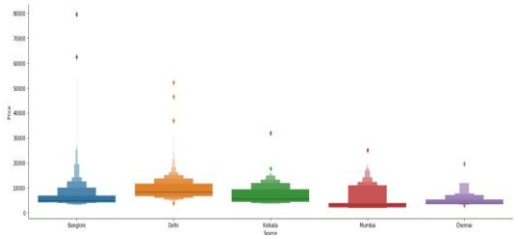
Figure 4.5: Updated table cont..



.Visualization

A graph, map, or other visual representation of data or information is known as data visualisation. It combines data and photographs to convey relationships. Machine learning facilitates the execution of analyses such as predictive analysis, which can then be presented as useful visualisations. This can be useful for detecting trends, corrupt data, outliers, and other things while exploring and getting to know a dataset.

Flight fare visualization w.r.t Airlines



Flight fare visualization w.r.t Cities

.Feature Selection

The feature selection is most important part of every machine learning project. Lot of features will needs to curse of dimensionality. By using heat map we remove similar (equally co-related) features because they both does the same job. The heatmap can be helpful to give a rough idea about the correlated features.

Figure 4.8: Heat map

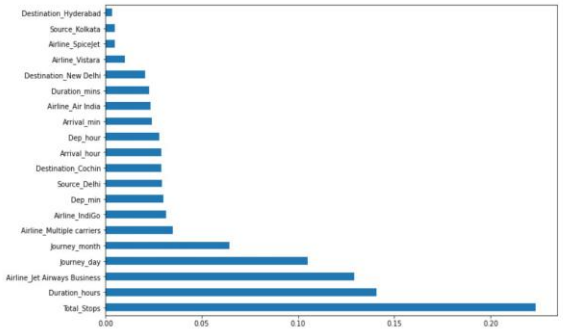


Figure 4.9: Feature Importance

4.0.4 ML Models and performance metrics

Random Forest

Hyperparameter Tuning

Evaluating your machine learning model is an essential part of any project. The model might give us satisfying results while using the evaluation metrics as accuracy but the same model might give poor results when we evaluate against other metrics such as logarithmic loss or such kinds of metrics. But we often use Accuracy to measure algorithmic performance. We going to work on different ml algorithms and compare the accuracy of each and every model. We are going to choose the best of all algorithms and train the model and generate a pickle file for deployment.

Deployment:The deployment of any machine learning model is nothing but we are preparing for the production where the web based applications, some enterprise softwares, and some of the APIs will take the trained model by giving new data and functional points and generating optimal predictions. As part of Deployment, we are saving our model in a pickle file. We use this pickle file in the flask app for using the model.

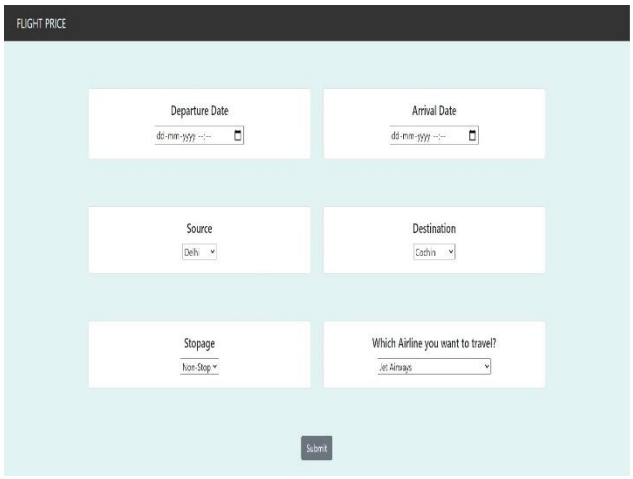
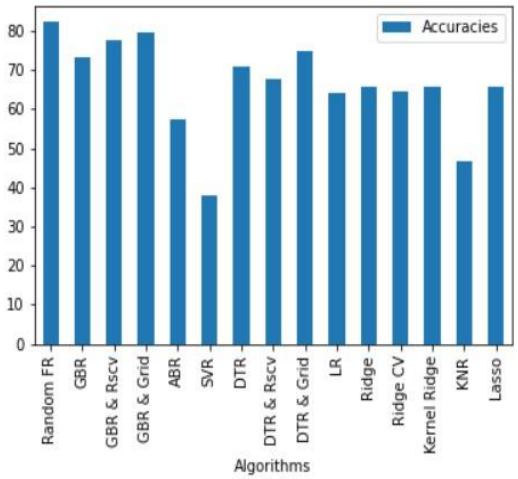


Figure 4.22: Home Page

Figure 4.23: Output Page

VI. RESULTS AND DISCUSSION

We have compared the accuracies of all the algorithms and found the best accuracy for the Random forest which is around 81 percent, As a result we have used random forest for our prediction.



Accuracy of Algorithms

Machine Learning Algorithms	Accuracy
Random Forest	0.811
Gradient Boosting	0.73
Gradient Boosting Random CV	0.77
Gradient Boosting Grid CV	0.79
Ada Boost	0.57
SVR Linear Kernal	0.37
SVR Rbf Kernal	0.02
Decision Tree	0.70
Decision Tree Random CV	0.67
Decision Tree Grid CV	0.74
Linear Regression	0.64
Ridge	0.65
Ridge CV	0.64
Kernal Ridge	0.65
K Neighbors Regressor	0.45
K Neighbors cross validation	0.55
Lasso	0.65

Algorithm vs Accuracy

VII. CONCLUSION

In this work we aim to gather airfare data from different airline corporation from different sources, study the data and prove that given the previous flight ticket prices it is possible to predict the new tickets fare. We applied multiple Machine Learning algorithms on the training dataset and chosen the best model which gives higher accuracy based on the R2 value. The other main factors for predicting the airfare are the accurate collection of data and selection of attributes from which we had the possibility to drew some useful conclusions. we understood and concluded which attributes have the most influence on the airfare prediction from the experiments that we conducted, which can be correlated either positively or negatively. In the next trend of advancement in technology, this work could be expanded to a greater extant to predict the airfare prices for the airplane ticket of the airline using real time data.

VII. FUTURE ENHANCEMENT:

Currently, we are only working with old and static Data sets. But in the future, we will try to work on time series data. The advantage of working in time series data is, based on the no of searches for the particular flight is directly proportional to the flight fare. If we work on time series this kind of prediction can also be done and the model becomes more accurate. In the near future, we would like to extend this work to predict the air prices for the entire industry. For this to be done, Additional experiments on larger flight fare data are essential. This model is explicitly for guiding the people to make the best airfare purchase in the best market.

REFERENCES:

1. S. Rajankar, O. Rajankar and N. Sarkharkar, "Predicting The Price Of A Flight Ticket With The Use Of Machine Learning Algorithms", INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH, vol. 8, no. 12, p. 3297, 2019.
2. Wang, Tianyi Pouyanfar, Samira Tian, Haiman Tao, Yudong Alonso, Miguel Luis, Steven Chen, Shu-Ching. (2019). A Framework for Airfare Price Prediction: A Machine Learning Approach. 200-207. 10.1109/IRI.2019.00041.
3. 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, pp. 1036-1039, doi: 10.23919/EUSIPCO.2017.8081365.
4. J. A. Abdella, N. Zaki and K. Shuaib, "Automatic Detection of Airline Ticket Price and Demand: A review," 2018 International Conference on Innovations in Information Technology (IIT), 2018, pp. 169-174, doi: 10.1109/INNOVATIONS.2018.8606022.