International Journal of Communication Networks and Information Security
2024, 16(4)
ISSN: 2073-607X,2076-0930
https://https://ijcnis.org/ Research Article

Localized Modeling for Airline Price Prediction Using K-Means and Decision Tree Ensemble

Mahek Upadhye[1*] , Chaitya Lakhani[1] , Rishab Pendam[1], Pranit Bari[2], Khushali Deulkar [2]

[1] Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
[2] Assistant Professor, Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

***Corresponding Author:** mahekupadhye123@gmail.com

ARTICLE INFO

Received: 29 Jul 2024
Accepted: 14 Sep 2024

ABSTRACT

Both travelers and airline companies rely on accurate prediction of flight prices, nevertheless it is difficult to train machine learning models using large-scale, current flight datasets due to computational inefficiency and risk of overfitting. This paper introduces a novel two-pronged approach that combines K-means clustering with decision trees for effective localized flight price forecasting. First, K-means clustering is employed to segment the flight data on the basis of shared characteristics into meaningful clusters thereby reducing data dimensionality and speeding up model training. Then individual Decision Tree models are built for each cluster separately. Finally, this technique emphasizes on closely related data attributes which may enhance predictive accuracy for given routes or types of flights. The proposed method solves the problem of calculation load when working with large datasets without sacrificing any details necessary for catching peculiarities in pricing in various categories of flights.

Keywords: Flight price prediction, K-Means clustering, Decision trees, Localized accuracy, Supervised learning after clustering

INTRODUCTION

Predicting flight prices is one of the continuous challenges that budget travelers and airlines face in a moving market. The landscape of the airline industry keeps changing, thus traditional methods of prediction do not work well with this kind of data like for instance real-time data streams which are enormous. However, there is a cost to using machine learning as a technique for predicting flight prices: it can be time-consuming or even suffer from overfitting if applied to large datasets. This article suggests an original method that combines traditional machine learning algorithms such as K-means clustering, k-Nearest Neighbors (KNN), and Decision Trees in order to make efficient and localized predictions about airfare.

The initial filter is K-Means clustering. It clusters data points into predetermined clusters (k) based on features relevant to flight pricing that they have in common. Examples of these characteristics include; Origin City, Destination City, Travel Time (direct/indirect), Days Until Departure, Number of Stops, Departure Time (morning/evening/etc). K-Means reduces the overall data size by grouping similar data points together. This greatly reduces training time needed for subsequent models.K-Means constructs clusters that describe specific travel situations. This has the advantage of allowing for more focused and perhaps more accurate predictions within each cluster as opposed to a global model predicting all flights at once.

After clustering, KNN is used to determine which cluster is best for price prediction. Prediction accomplished by employing Decision Tree regressors, which may be useful in forecasting the appropriate price in a variety of situations by potentially spotting intricate relationships between data sets. This method allows for the identification of subtle pricing patterns across a variety of flight categorizations, even as the computational challenges associated with processing large datasets are being addressed.

The three main components of the suggested method's general strategy are preprocessing data using K-Means, classifying clusters using KNN, and estimating final prices using regression trees. Taken together, these steps provide a comprehensive approach to solving flight fare prediction problems. Predictions may become more accurate as a result, especially for non-linear price trends.

Overall Strategy:

1. Divide flights with similar attributes into groups by applying the K-Means clustering algorithm to the data.

2. K-mean Nearest neighbors classification is used to predict a suitable cluster out of k nearest neighbors. Finally, a decision tree regressor is built, which can serve as an identifier of price trends in the data sets of the clusters.

This strategic measure not only reduces the data dimension but also fast tracks the training process for subsequent Decision Tree models. The cluster-based approach, instead of the overall model, focuses on the data reduction and uses a particular model to each cluster to successfully capture the subtle pricing patterns.

LITERATURE REVIEW

In [1] a holistic machine learning approach is proposed for airfare prediction. The model looks at several factors besides historical prices, such as flight seasonality, demand, and social media sentiment. This approach focuses on more accurate predictions than traditional ones. The authors run a comparative study with baseline models to prove their model’s effectiveness. This particular study emphasizes the potential of machine learning for airfare prediction using various influencing factors.

[2] Investigates the use cases of machine learning for forecasting flight fares. The paper explores historical flight data and leverages a machine learning technique, specifically Random Forest, to identify patterns and relationships influencing flight ticket prices.

In [3] paper, the authors take a deep dive into machine learning algorithms and their application in predicting flight fares. Rather than sticking to conventional methods, the research focuses on identifying the driving forces behind ticket prices. Random Forest and hyperparameter tuning emerge as essential techniques in this study, offering a more refined approach to cost optimization. The researchers emphasize the importance of computational modeling in uncovering these insights, aiming for better pricing strategies.

The study in [4] explores airfare variability in the Indian market using Machine Learning and AutoML techniques for prediction. It analyzes factors like flight duration, destination, and events, employing visualization methods such as scatter plots and ggplots. Random Forest Regressor and Randomized Search CV achieved the best prediction accuracy.

The author in [5] implemented a Deep Neural Network that simulates brain functions. The dataset was preprocessed using Min-Max normalization to enhance performance, and the Randomized Search CV algorithm was applied for hyperparameter tuning. Additionally, univariate, bivariate, and correlation analyses were used to visualize the dataset's features.

[6] explores factors affecting airfare, such as flight schedule, destination, and special occasions. It analyzes three datasets from Kaggle, MakeMyTrip and Data World applying machine learning models like KNN, Linear, Lasso, Ridge, and Random Forest regression. Random Forest Regressor showed the highest accuracy in predicting ticket prices, helping identify key pricing factors for a predictive fare model.

In [7], the authors have applied the K-Nearest Neighbors (KNN) technique to estimate flight prices using machine learning methods. It compares airfare variations across different days, weekends, and times of day. Regression analysis is then performed to predict flight prices, aiming to identify the optimal purchasing times based on these factors.

The [8] study takes a more data-driven path, concentrating on airfare prediction using 1,814 flights operated by Aegean Airlines between Thessaloniki and Stuttgart. By training various machine learning models on this dataset, the authors achieved an impressive accuracy of nearly 88%. This paper evaluates eight different models, providing readers with information on which algorithms perform best under specific circumstances. The comprehensive comparison of the models is an important contribution to the field of airfare prediction.

In [9], the authors address the dynamic nature of flight pricing using machine learning.Key gaps include a lack of algorithm comparison, missing evaluation metrics, and no mention of real-time data integration. The absence of model validation or algorithm comparison limits its impact.

[10] developed a system to help consumers decide the best time to buy airline tickets. Their model used partial least squares (PLS) regression to predict optimal purchase timing. They applied techniques like feature extraction, creating time-lagged data, and building regression models to find the most effective approach. PLS regression stood out as the best method, thanks to its ability to filter out irrelevant and highly correlated factors.

The paper [11] provides a comparative analysis of two popular machine learning algorithms for predicting flight ticket prices. The study demonstrates that the Random Forest Regressor outperforms the Decision Tree Regressor in terms of accuracy and robustness for airfare forecasting. The authors highlight the effectiveness of Random Forest due to its ensemble nature, which reduces overfitting and improves predictive performance.

The study conducted in [12], focuses on enhancing flight price prediction through machine learning integration. The research shows that incorporating these techniques into existing systems can significantly boost both the accuracy and efficiency of price forecasts. By merging advanced algorithms with current platforms, the study highlights tangible improvements in prediction performance. This advancement is crucial for developing more responsive and dependable systems in airfare forecasting, benefiting consumers and the travel industry.

The authors of [13] introduced a method that combines Self-Organizing Maps (SOM) with Least Squares Support Vector Machines (LSSVM) for predicting flight prices. The approach involves preprocessing flight data, optimizing feature selection with Enhanced Harmony Optimization (EHO), and achieves a high accuracy rate of 97.41%, surpassing existing methods.

METHODOLOGY

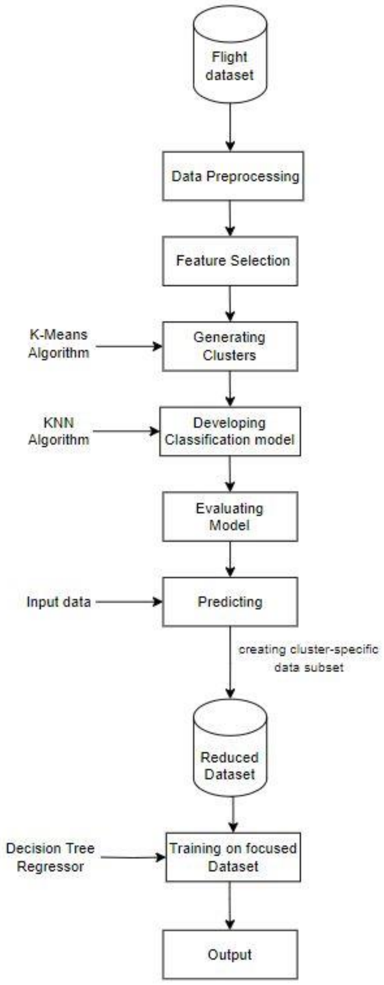
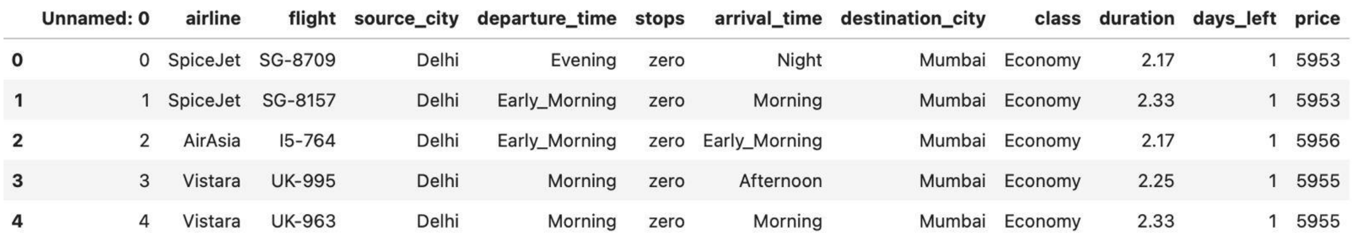


Figure 1. Proposed Methodology

Description of Dataset

The dataset [16] analyzed in this paper includes 300,261 individual flight booking options collected over fifty days. The dataset contains specific information on each flight, a specific booking inclusive of many variables that influence the fare prices. Key features of the dataset include the airline name, the origin and destination cities, the layover details as well as the times of departure and arrival. Furthermore, it keeps track of the total duration of each flight and the number of stops, both of which are the main factors that are influencing the fare price production. The target feature in this dataset is the flight fare, which the models are expected to predict.



	Unnamed: 0	airline	flight	source_city	departure_time	stops	arrival_time	destination_city	class	duration	days_left	price
0	0	SpiceJet	SG-8709	Delhi	Evening	zero	Night	Mumbai	Economy	2.17	1	5953
1	1	SpiceJet	SG-8157	Delhi	Early_Morning	zero	Morning	Mumbai	Economy	2.33	1	5953
2	2	AirAsia	I5-764	Delhi	Early_Morning	zero	Early_Morning	Mumbai	Economy	2.17	1	5956
3	3	Vistara	UK-995	Delhi	Morning	zero	Afternoon	Mumbai	Economy	2.25	1	5955
4	4	Vistara	UK-963	Delhi	Morning	zero	Morning	Mumbai	Economy	2.33	1	5955

Figure 2. Snapshot of Dataset

Data Preprocessing

This study begins with comprehensive data preprocessing and the normalization of raw data. Data preparation includes first loading the flight dataset to pandas DataFrame which subsequently serves as the core for advancing analysis. The focus of the analysis centers around the distribution of costs between source and destination pairs and the fees collected by different airlines and flight combinations. This examination produces many advantageous insights about outliers as well as pricing tendencies towards specific routes and carriers. Following this initial assessment

Following this, a detailed feature selection process is undertaken. Irrelevant features are identified and eliminated, this has reduced the dimensionality of the dataset. For example, the arrival time was deemed unnecessary due to the presence of departure time and flight duration, which were sufficient in capturing the temporal aspects relevant to pricing. This step is crucial for enhancing the efficiency of subsequent model training procedures.

The features retained are analyzed for their unique values, which are the key information for the feature engineering decisions to be made. The visualizations are then used as tools for spotting outliers, overviewing skewness, and distributional imbalances which are all potential causes of the model's performance.

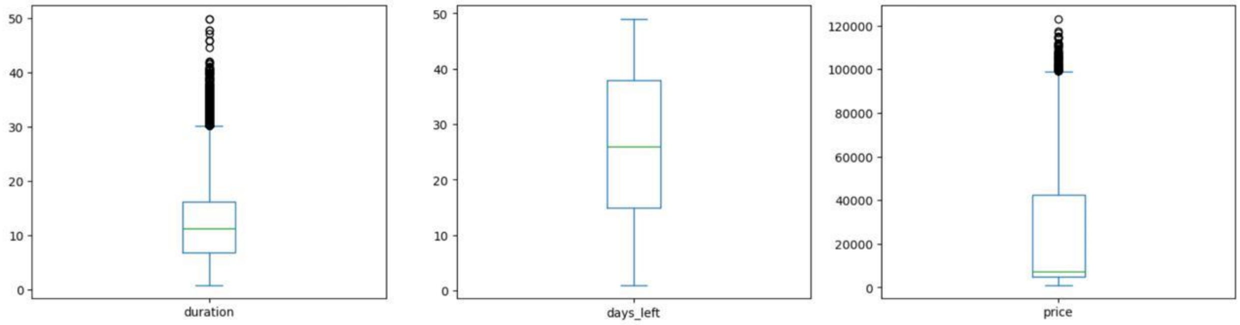


Figure 3. Distribution of data

A variety of techniques were used to prepare the data for the machine learning modeling, during preprocessing phase. To start with, numerical variables such as the number of stops were initially stored as strings and were converted into integers to be compatible with traditional algorithms. Followed by, one-hot encoding that is used to create dummy variables for categorical attributes airline such as source city, departure time ('Early Morning', 'Morning', 'Afternoon', 'Evening', 'Night', 'Late Night'), destination city. In the end, numerical attributes such as duration and days left were scaled through StandardScaler to put them in a similar range. This normalization step helped mitigate the risk of features with larger scales dominating the training process and enabled the models to learn from the relative importance of each attribute. These preprocessing steps transformed the data into a form appropriate for subsequent machine learning analysis.

This preprocessing approach establishes a robust foundation for the subsequent stages of the study, enhancing the potential for accurate and reliable flight price predictions.

Model Development

1. K-Means Clustering

K-means clustering in the study is systematically done through optimization of cluster count and splitting of flight data into meaningful groups. This process is executed in three distinct steps:

Step 1: The elbow method is used to determine the optimal number of clusters. In this technique, a range of k values (number of clusters) are iterated and the Sum of Squared Errors (SSE) for each k is computed. SSE is defined as:

$$SSE = \sum_{i=1}^{n} ||x_i - c_j||^2$$

Where x_i individual data points and c_j represents the centroid of cluster to which x_i belongs. Looking at the point where the bending or elbow occurs in relation to SSE versus k plot will enable one identify how many clusters should be created; beyond that adding more clusters will not lead to significant reduction in SSE. This methodology helps deal with overfitting while capturing some inherent groupings within data.

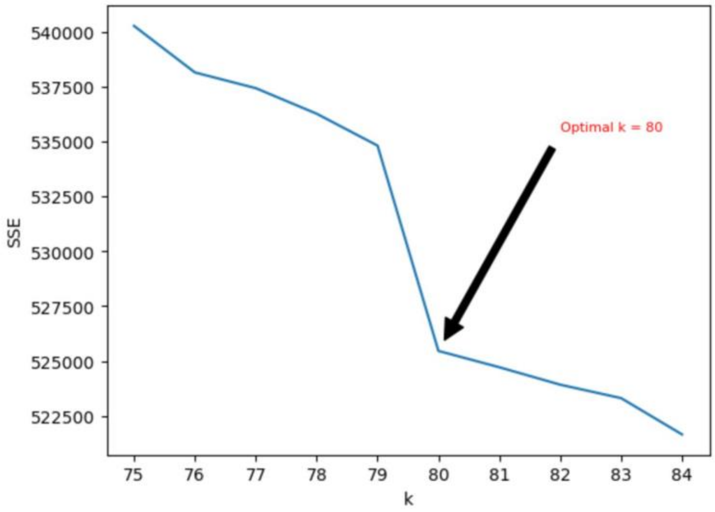


Figure 4. Elbow Plot

Step 2: The next step consists of K-means clustering after determining an optimum number of clusters. This algorithm groups flights based on the similarity between various attributes such as origin, destination, travel time, days until departure, number of stops, and departure time among others. Similar pricing patterns may be found in these clusters by grouping them.

Step 3: Individual data points (flights) are given cluster labels according to how close they are to cluster centroids in the feature space. Flights are assigned to the cluster with the closest centroid based on feature similarity. Each centroid is a cluster's center point. For every flight, this procedure creates a new feature that indicates cluster membership, making further analysis and modeling easier.

Using this clustering technique paves the way for more complex price prediction models that might be able to identify regional pricing patterns among comparable flight classes.

2. K-Nearest Neighbor(KNN)

The K-Nearest Neighbors (KNN) algorithm is a classification method widely used to assign a class to a test instance based on the most frequent class among its 'k' nearest neighbors. The main idea behind KNN is that a data point is classified by examining the classes of its nearest data points in the feature space.

KNN calculates the distance between the test instance and all instances in the training dataset to determine the nearest neighbors. The Euclidean distance, a typical measure, is determined as follows:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{i,k} - x_{j,k})^2}$$

where $x_{i,k}$ and $x_{j,k}$ are the k-th feature values of instances x_i and x_j respectively.

For a given test point x, the algorithm computes distances to all training points and selects the k closest points based on these distances.

The test instance is classified based on the majority class among the k nearest neighbors. The class c^* is assigned as:

$$c^* = \text{mode}(c_1, c_2, \ldots, c_k)$$

where $\{c_1, c_2, ..., c_k\}$ are the classes of the nearest neighbors. The following steps are undertaken in training the KNN classifier on the flight dataset, which has been appended with cluster membership labels:

Step 1: A grid search is performed to discover the best number of neighbors (k). This consists of assessing the KNN classifier's results through a range of k values using cross-validation. Cross-validation creates as many as several different folds of the data and then averages the model's result over the folds to hold overfitting in a bay. The grid search chooses the k with the highest model's accuracy or any other performance metrics.

Step 2: With the determined optimal k, the KNN classifier is trained using the training data. The model learns to label instances by the most common class of their closest neighbors. The trained KNN model is then tested with the test set to assess its classification accuracy. The model got an accuracy score of about 96.6%, which shows its ability to classify test instances based on their nearest neighbors.

3. Decision Tree Training

A Decision Tree Regressor is a type of machine learning algorithm that can predict continuous values in a given data set. It utilizes the idea of dividing information by different attributes, which resembles a tree structure of decisions. Each inner tree node is a choice that depends on the value of a specific feature while each leaf node is the expected output value. The algorithm finds the best way to split the data at each node by considering factors such as variance reduction or mean squared error (MSE). The objective is to establish the quickest possible subsets in the end. This process continues until a stopping criterion, like a maximum tree depth or a minimum number of samples per leaf, is met.

The research evolves from a one-model method to a more advanced cluster-specific decision tree training technique. This novel method of analysis, which involves K-Means clustering, is the tool with which the predictions of flight prices can be more accurate.

Step 1:The clusters that have been identified by the pre-trained KNN classifier are presented as separate data subsets that consist of the flights that only belong to that cluster. Through the division of the data, it will be possible to conduct a dedicated evaluation of the characteristics of the flights within each group.

Step 2: A separate Decision Tree Regressor is trained for each of the predicted clusters. In this way, every individual model is given the opportunity to concentrate on detecting the patterns in a more unified group of flights, thereby capturing more intricate connections between features and prices within each cluster.

This technique of training per-cluster has multiple benefits. For instance, decision tree models may identify more specific relationships between the features and flight prices by restricting themselves to the data within a certain cluster.

Moreover, in a cluster where domestic short-haul flights are mainly found, one could expect a pricing pattern to differ from the other where the majority of routes are international or long-haul. By incorporating the discovered structure, the method aims to produce better results when compared to the use of the single model. Hence, category-specific models in cluster-specific decision tree training are able to generate more accurate and context-aware price predictions for a wide variety of flight scenarios.

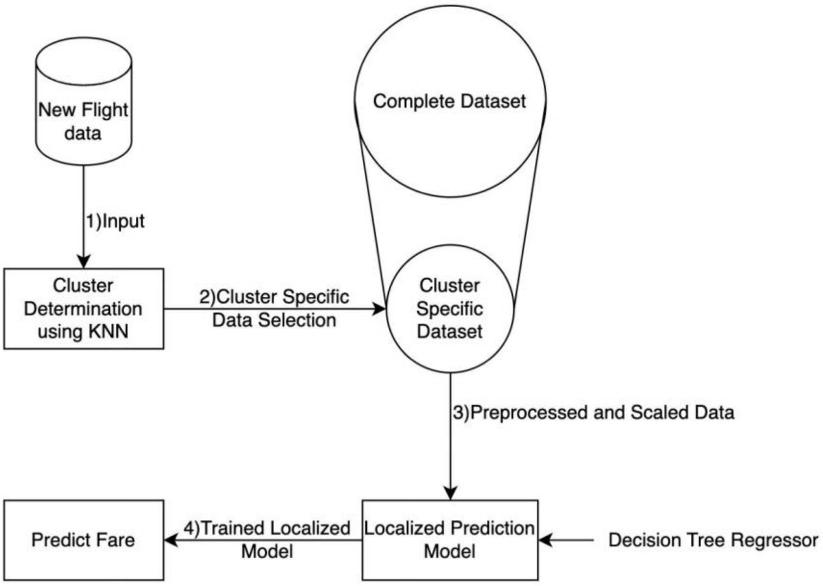

New Flight data
1)Input
Cluster Determination using KNN
2)Cluster Specific Data Selection
Complete Dataset
Cluster Specific Dataset
3)Preprocessed and Scaled Data
Predict Fare
4)Trained Localized Model
Localized Prediction Model
Decision Tree Regressor

Figure 5. Prediction Mechanism

Prediction on flight data

Flight data is the input of the prediction function. Thereafter, the pre-trained KNN classifier is used to predict which cluster the flight belongs to by analyzing its features. The preprocessed and scaled data is then split into training and testing sets (X-train, X-test, y-train, y-test) using a train-test split. Finally, for each cluster, a decision tree regressor is trained in order to find their prices. The price of a flight would be determined by this model, built specifically for this cluster.

RESULTS AND DISCUSSION

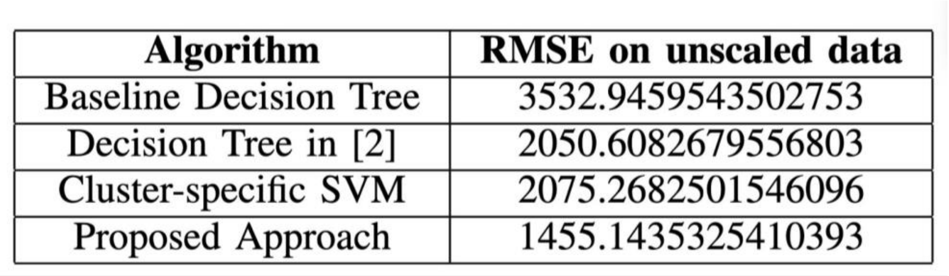
Metrics, such as root mean square error (RMSE) and mean absolute error (MAE), are used in evaluating performance across models, both global and localized cluster specific models, in terms of predictions made against test set values.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where,

- y_i : The actual value for the *i*-th data point.
- $\hat{y}_i$: The predicted value for the *i*-th data point.
- n: The total number of data points.

Table 1. COMPARISON AMONG VARIOUS TECHNIQUES



Algorithm	RMSE on unscaled data
Baseline Decision Tree	3532.9459543502753
Decision Tree in [2]	2050.6082679556803
Cluster-specific SVM	2075.2682501546096
Proposed Approach	1455.1435325410393

Initially, the entire flight dataset was utilized as input to Decision Tree regressor which produced a high RMSE of 3532.9459543502753 that yielded computational cost and suboptimal model performance globally. Similarly, Support Vector Machine (SVM) regressor was explored as an alternative to the baseline Decision Tree model. The SVM regressor was trained on the full dataset but results showed a very high RMSE meaning that the SVM model failed to attain the desired accuracy compared with the decision tree regressor. To overcome limitations of global models, cluster-specific approach was employed. The dataset was divided into clusters with similar properties using K-Means clustering so that pricing predictions could be more localized and possibly more accurate. Thereafter, SVM regressor was used on each data subset based on its cluster resulting in an RMSE of 2075.2682501546096. Finally, this paper implemented a proposed approach that includes K-Means clustering, KNN classification and Decision Tree regressors. By employing such localized modeling technique, this method managed to achieve better performance, reducing RMSE to 1455.1435325410393, compared to all other methods before it.

CONCLUSION

This study introduces a new method for predicting flight prices by combining K-Means clustering, K-Nearest Neighbors (KNN), and Decision Trees. The process starts with K-Means clustering, which simplifies the data and makes it easier to handle, reducing both the time and computational effort required. Next, KNN classifies these simplified data clusters, which helps improve the accuracy of the predictions. Finally, Decision Trees work with these refined clusters to deliver precise and relevant predictions. This approach is more than just a technical improvement; it offers practical benefits. It makes the prediction process faster and more energy-efficient than traditional methods, which can be complex and resource-intensive. By handling large datasets more effectively, it provides a more accurate and practical way to forecast flight prices. This means travelers can make better decisions about their bookings, potentially saving money and time. Overall, this method not only enhances prediction accuracy but also makes the process more manageable, offering a smarter and more effective solution for predicting flight prices.

FUTURE SCOPE

The present methodology of K-Means clustering and Decision Trees are strong enough to provide a robust baseline for predicting flight prices with good absolute price predictions, although there is still plenty of improvement that can be achieved with a modest extension. An essential next step is to forecast how the prices of a particular flight would evolve over time. With this prediction, the traveler can be more informed about when to book a flight to maximize the savings. Deep learning architectures, e.g., RNNs, are natural for sequential data, such as historical flight prices. The inclusion of historical data and even other features such as the booking window and seasonality may enable RNNs to learn the complex patterns in the historical data and predict the future trends of the prices. Users could input the relevant flight parameters, see a current flight price prediction for their parameters, and then find out how the price would change in the coming weeks or months based on historical rates of change and the factors that influence price changes. In addition, the techniques of hyperparameter tuning for both the K-Means clustering and Decision Trees, which were created in the previous K-Means clusters, may also be useful in improving the performance of the K-Means clustering and Decision Trees, respectively, therefore, increasing the overall accuracy and the absolute price predictions will be improved.

REFERENCES

[1] T. Kalampokas, K. Tziridis, N. Kalampokas, A. Nikolaou, E. Vrochidou, and G. A. Papakostas, "A Holistic Approach on Airfare Price Prediction Using Machine Learning Techniques," *IEEE Access*, vol. 11, pp. 46627–46643, 2023, doi: 10.1109/ACCESS.2023.3274669.

[2] K. D. V. N. Vaishnavi, L. H. Bindu, M. Satwika, K. U. Lakshmi, M. Harini, and N. Ashok, "Flight Fare Prediction Using Machine Learning," *EPRA International Journal of Research and Development (IJRD)*, vol. 8, no. 10, Oct. 2023, doi: 10.36713/epra14763.

[3] N. Alapati *et al.*, "Prediction of Flight-fare using Machine Learning," in *2022 International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP)*, Uttarakhand, India, 2022, pp. 134–138, doi: 10.1109/ICFIRTP56122.2022.10059429.

[4] A. Karambelkar, P. Mamania, and V. Chunekar, "Analysis of Flight Fare Detection Using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 11, no. 11, Nov. 2022.

[5] G. Ratnakanth, "Prediction of Flight Fare Using Deep Learning Techniques," in *2022 International Conference on Computing, Communication and Power Technology (IC3P)*, 2022, pp. 308–313, doi: 10.1109/IC3P52835.2022.00071.

[6] R. R. Subramanian, M. S. Murali, B. Deepak, P. Deepak, H. N. Reddy, and R. R. Sudharsan, "Airline Fare Prediction Using Machine Learning Algorithms," in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, pp. 877–884, doi: 10.1109/ICSSIT53264.2022.9716563.

[7] S. N. Prasath, S. Kumar M, and S. Eliyas, "A Prediction of Flight Fare Using K-Nearest Neighbors," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1347–1351, doi: 10.1109/ICACITE53722.2022.9823876.

[8] Z. Zhao, J. You, G. Gan, X. Li, and J. Ding, "Civil Airline Fare Prediction with a Multi-attribute Dual-stage Attention Mechanism," *Applied Intelligence*, vol. 52, no. 5, pp. 5047–5062, 2022.

[9] M. Lu, Y. Zhang, and C. Lu, "Approach for Dynamic Flight Pricing Based on Strategy Learning," *Journal of Electronics and Information Technology*, vol. 43, no. 4, pp. 1022–1028, 2021.

[10] K. Tziridis, T. Kalampokas, G. A. Papakostas, and K. I. Diamantaras, "Airfare Prices Prediction Using Machine Learning Techniques," in *2017 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017, pp. 1036–1039, doi: 10.23919/EUSIPCO.2017.8081365.

[11] S. Gupta and N. Gupta, "Flight Fare Prediction Using Machine Learning," *Journal of Computational Science and Engineering*, 2023.

[12] W. Groves and M. Gini, "An Agent for Optimizing Airline Ticket Purchasing," in *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, May 2013.

[13] N. S. S. V. S. Rao and S. J. J. Thangaraj, "Flight Ticket Prediction Using Random Forest Regressor Compared with Decision Tree Regressor," in *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Apr. 2023, pp. 1–5, doi: 10.1109/ICONSTEM.2023.9673263.

[14] S. M. Joshitta, M. P. Sunil, A. Bodhankar, C. Sreedevi, and R. Khanna, "The Integration of Machine Learning Technique with the Existing System to Predict the Flight Prices," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, May 2023, pp. 398–402.

[15] R. Umanesan, S. Raha, B. Ganguly, A. Thangam, and A. Bhongade, "A Novel Approach to Predict Flight Fares in India Using SOM-LSSVM Model Approach," in *Proceedings of the International Conference on Electronics, Communication, Computing and Control Technology (ICECCC 2024)*, Bangalore, India, 2024, doi: 10.1109/ICECCC61767.2024.10593942.

[16] S.Bathwal, "Flight Price Prediction," 2021, available at: https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction.

[17] H. Korkmaz, "Prediction of Airline Ticket Price Using Machine Learning Method," *Journal of Transportation and Logistics*, vol. 9, no. 2, pp. 1–14, 2024, doi: 10.26650/JTL.2024.1486696.

[18] B. Burger and M. Fuchs, "Dynamic Pricing – A Future Airline Business Model," *Journal of Revenue and Pricing Management*, vol. 4, no. 1, pp. 39–53, 2005.