AIRLINES PASSENGER SATISFACTION SYSTEM

Kundrapu Vineetha
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetam
Bengaluru, India
BL.EN.U4CSE21106@bl.students.amrit.edu

Mettukuru Tharun Reddy
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetam,
Bengaluru, India
BL.EN.U4CSE21123@bl.students.amrita.edu

Narisetty Prathima
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetam
Bengaluru, India
BL.EN.U4CSE21132@bl.students.amrita.edu

Md Jaffer Ali
Department of Computer Science and Engineering
Amrita School of Computing,
Amrita Vishwa Vidyapeetam,
Bengaluru, India
BL.EN.U4CSE21124@bl.students.amrta.edu

***Abstract*- Air travel is an integral part of modern life, with millions of passengers taking to the skies every day. Since passenger satisfaction directly affects consumer loyalty and company performance, airlines all over the world now prioritize ensuring passenger satisfaction.**
The main objective of this study is to analyze the performance of various models. Supervised models like Logistic Regression with Elastic Net, Naive Bayes, Decision Trees, Random Forest, Neural Networks to predict and understand passenger satisfaction levels. Additionally, unsupervised methods such as K Means, DBSCAN, and Hierarchical Clustering aid in uncovering hidden patterns within the data. Feature selection techniques like Chi-Square and Wrapper Methods further refine the model's understanding, contributing to a holistic approach aimed at enhancing the airline passenger experience. This helps to improve experience and optimizing operations based on the feedback.

Keywords- Random Forest, Decision Tree, Airlines, Naïve Bayes.

I. Introduction

The principal aim is to enhance the overall enjoyment of air travel. In the current era of intense airline competition and growing air travel development, passenger satisfaction is crucial. It seeks to offer guests a sophisticated, customized, and enjoyable experience. Airlines need to strike a balance between offering top-notch customer service, being cost-effective, and being conveniently located. In addition to being essential for customer retention, passenger satisfaction also affects an airline's profitability and market share. It serves as a guide for those engaged in enhancing the experience of travellers and attaining the airline industry's sustained expansion.

This research helps enables airlines to focus on enhancements that have a direct effect on the customer experience. Additionally, it led to improvements in operations, simplifying procedures to improve passenger satisfaction and efficiency during travel. The ability to track changes in customer satisfaction over time enables airlines to modify their plans for growth and maintain high service standards. It makes it easier to include direct feedback from customers in initiatives for improvement, enabling in-the-moment modifications and immediate solutions to issues.

II. Literature Review

Bhargay [1] indicates an in-depth look using machine learning techniques of passenger happiness in the airline sector. Demonstrates the use of K-nearest neighbour (KNN) and a Novel Hybrid Random Forest model for prediction.
The findings imply that the Hybrid Random Forest performs more accurately than KNN.It is suggested that future elfort improve accuracy and shorten computation times. Keerhy A S [2] focuses on how important passenger pleasure is on airlines. It shows the value of passenger feedback data and looks at numerous studies examining consumer happiness across various airlines. The study uses classification techniques like KNN, Logistic Regression. Random Forest. Decision Tree, SVM, and Naive Bayes to find factors influencing passenger happiness. Li Yan-wei [3]The study emphasizes the value of abandoning traditional methods and the use of SEM to system analysis and rationality evaluation of customer satisfaction indices. In addition to providing a helpful SEM-based model with eight reflective variables, the study carefully examines the factors influencing customer satisfaction in the airline industry. The survey highlights how much more effective SEM can be when compared to traditional techniques. Kumar Pramod Panda [4] investigates the sentiment analysis of customer satisfaction in airline tweets using machine learning. A wide range of techniques have been examined, such as logistic regression, AdaBoost, decision trees, random forests, support vector machines, K-nearest neighbors, and Gaussian Naive Bayes. An efficient ensemble classifier is recommended by the study's result in order to handle customer satisfaction more successfully in the

future. Nikola Vojtek[5] emphasizes the significance of precise data collection and analysis in order to improve flying services. The assessment places emphasis on feedback channels, including publicly available websites like TripAdvisor and software-as-a-service solutions employed by airlines. It references studies on loyalty programs, the attributes of first-rate service, and the impact of social media on consumer ratings. The findings underscore the need for impartial feedback systems and promote further investigation into innovative approaches such as blockchain technology. Usha, Dr. P. [7] According to the study, the most crucial elements influencing passenger satisfaction were customer service, food and drink, in-flight amenities, and on-time performance. According to the study's findings, Air India can raise customer satisfaction levels by concentrating on enhancing its check-in procedures, in-flight amenities, particularly the food and drink options, on-time performance, and customer service. R. Archana [8] According to the survey, the three factors that most significantly influence passenger satisfaction are safety, seat comfort, and in-flight entertainment. The study's conclusions suggest that Indian Airlines should prioritize improving in-flight entertainment, seat comfort, and safety standards in order to boost customer satisfaction.Chang [9] The three factors that have the biggest effects on passenger satisfaction were found to be safety, seat comfort, and in-flight entertainment. In 2002, the Journal of Operational Research published the article.As part of the study, 300 domestic airline passengers in Taiwan were surveyed to learn more about their satisfaction with various aspects of the services offered by the airlines. Customers were pleased with airlines that had a good safety record. Ostrowski [10] In order to gather data on the passengers' satisfaction with different aspects of the airline services, 300 passengers from two major US airlines were surveyed. Punctual performance Airlines that did well in this area received higher satisfaction ratings from passengers. Consumers expressed satisfaction with the airline's overall quality of service, the food and beverages provided, the state of the aircraft, and the manner in which airline employees handled them. Price: Passenger satisfaction was higher on airlines that offered good value for the money. The study also found a positive correlation between passenger satisfaction and customer loyalty. Mehta[11] validate their hypotheses, the writers gathered data from an Indian airline passenger survey consisting of 300 respondents. The results of the study show that the caliber of services offered has a big impact on customer satisfaction and loyalty. Furthermore, the scientists found that passenger pleasure has a substantial positive impact on loyalty. The results of the Tsaur [12] study showed that customer satisfaction and loyalty are highly impacted by the caliber of the services offered. The writers also found that customer satisfaction greatly increases loyalty. J. Li [13] This study offers a thorough analysis of the literature on airline passenger satisfaction. A total of 109 publications met the inclusion criteria that the authors were looking for. According to the analysis, service quality, cost, convenience, and reliability are the factors that have the biggest impacts on airline passenger satisfaction. The researchers also found that satisfied consumers are more likely to recommend the airline to others and use it again. A loyalty and satisfaction model for airline passengers is developed and assessed by Chen, C. C. [14]. The model considers factors including quality, affordability, convenience, and dependability of service. Data from a survey of 400 Chinese airline passengers was used to evaluate the model. The results of the study showed that the model fits the data well and that each of its elements greatly raises consumer satisfaction and loyalty. A study on customer satisfaction with Sivasundaram. VS [15] found that the airline offers pleasant and friendly crew, comfortable seating, excellent catering, and entertainment options. The study emphasizes the strong correlation between service quality and customer satisfaction, underscoring the crucial role that meeting customers' needs for top-notch services plays in the airline industry.

III. METHODOLOGY

Dataset-

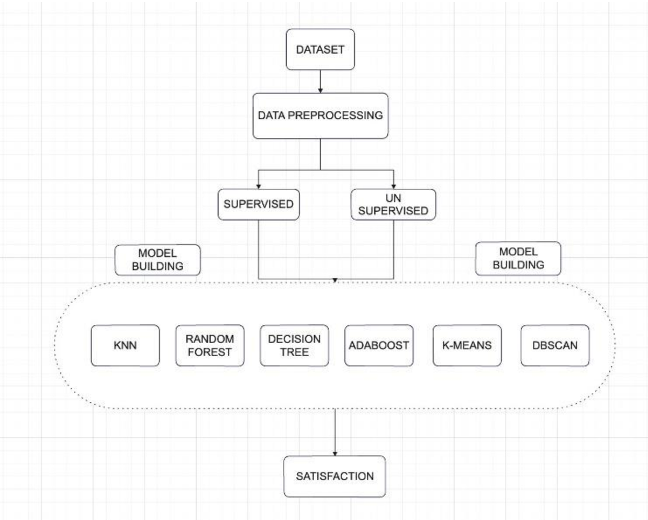
Creating a passenger satisfaction system that works well for airlines is difficult. As it relates to our dataset, it includes data on airline passengers and includes columns like "Gender," "Customer Type," "Age," "Type of Travel," "Class," and various metrics related to customer satisfaction. With 103,904 entries, this dataset is a useful tool for analysis and modeling since it captures the subtleties of passenger experiences.

It explores topics such as "Wifi in-flight," "On-board services," "Cleanliness," "Departure Delay in Minutes," and "Arrival Delay in Minutes" before putting an overall "satisfaction" label at the end. The intention is to improve the caliber of airline services by comprehending and forecasting passenger satisfaction levels. containing a variety of responses, from "neutral or dissatisfied" to "satisfied."

A. Data Preprocessing-

To make sure the dataset was appropriate for various models, it was systematically transformed. Initial steps involved importing training and testing data, then eliminating 'Unnamed: 0' and 'id', which were deemed unnecessary. After assessment, it was discovered that the target variable's imbalance was evenly distributed. Then, missing data was handled, with a particular emphasis on the mode imputation forcategoricalfeatures('Class,'Gender,'Customer_Type,'Type_of_Travel,'and'Arrival_Delay_in_Minutes') and mean imputation for numerical features. Through the use of exploratory data analysis (EDA), relationships between different features and passenger satisfaction were visualized. To make categorical variables compatible with machine learning algorithms, they were label-encoded. Interquartile Range (IQR) was used to identify and eliminate outliers. Using a heatmap, the relationship between the features was investigated. Finally, the chi-squared statistical test was used to select features, identifying the top 10 relevant features.

Fig.1 Generic Architecture



SUPERVISED:

A. Logistic Regression

A common binary classification method for determining the probability that an instance will belong to a specific class is logistic regression. Logistic regression model using hyperparameters that enhance the model's capacity for generalization, such as an elastic net penalty, a regularization mixing parameter (l1_ratio) set to 0.5, and the'saga' solver. Using a custom run_model function, the algorithm is applied to training and testing datasets separately.Based on the given hyperparameters, logistic regression—which is preferred for binary tasks—balances the L1 and L2 regularization penalties.

B. Naive Bayes

Based on the Bayes theorem, the Naive Bayes algorithm is a probabilistic classification method. Computing probabilities gets simpler because features are believed to be conditionally independent given the class label. Given the limited number of configurable parameters of Gaussian Naive Bayes, the model is initialized with no hyperparameters supplied (an empty dictionary called params_nb). Next, using various training and testing datasets, the model is trained and evaluated using the custom run_model function.

C. K-Nearest Neighbor Classifier

The majority class or average of the k nearest neighbors in the feature space is used by K-Nearest Neighbors (KNN) to assign data points. It is applied to both regression and classification tasks. N_neighbors, kd_tree (neighbor computation algorithm), and n_jobs (number of parallel jobs) are examples of hyperparameters for the KNN model. The KNN model is then trained and assessed on the provided datasets using the run_model function, producing a classification report, accuracy, and ROC Area under the Curve.

D. Decision Tree

A decision tree is a versatile machine learning technique that divides data systematically based on feature values to produce a decision structure that resembles a tree. It is used for regression and classification. It is widely recognized for being interpretable and offers insights into the decision-making process. In this specific implementation, a Decision Tree classifier is instantiated with hyperparameters, the maximum tree depth (max_depth) is set to 12, and each split's square root of features (max_features) is taken into account.

E. Random Forest

During training, a large number of decision trees are constructed using an ensemble learning technique called Random Forest, which yields the mode of the classes for classification tasks or the average prediction for regression problems. returns the initial values of a Random Forest model's hyperparameters. The hyperparameters are: 100 is the total number of trees in the forest; 1 is the minimum sample per leaf; 2 is the minimum sample required to divide an internal node; and a random seed is used for repeatability. The maximum depth of 16 is for the trees.

F. Neural Network (Multilayer Perceptron)

The structure of a neural network, and specifically a Multilayer Perceptron (MLP), a family of artificial neural networks, is defined by its input layer, one or more hidden layers, and output layer. Backpropagation is a supervised learning technique used to train MLPs. Using specific hyperparameters, the code provided instantiates an MLPClassifier from scikit-learn. The neural network comprises three hidden layers, each with thirty nodes, and employs the logistic activation function.

G. Extreme Gradient Boosting

XGBoost, or Extreme Gradient Boosting, is a potent technique that works well for both regression and classification issues. It develops a series of weak learners—typically decision trees—increasingly and optimizes their combination to increase prediction accuracy. The code snippet above sets the number of estimators to 500 and caps the maximum tree depth at 16 when configuring an XGBClassifier with specific hyperparameters.

H. Adaptive Gradient Boosting

Adaptive gradient boosting, or AdaBoost, is an ensemble learning technique that combines the predictions of weak learners—typically decision trees—to create a dependable and accurate predictive model. AdaBoost changes the weights given to training instances during the learning process, giving those that were previously wrongly labeled by weak learners more weight. Using specific hyperparameters, the provided code instantiates an AdaBoostClassifier from scikit-learn.

I. SVM

Support Vector Machine (SVM) is a supervised machine learning technique used for regression and classification tasks. The given code imports the Support Vector Classifier (SVC) from the scikit-learn SVM module. Using a support vector machine (SVM), the optimal hyperplane for classifying data points is found. After instantiating the clf (classifier) object, the run_model function is used to train and evaluate the SVM model on a dataset with attributes related to passenger satisfaction.

HYPERPARAMETER TUNING

The act of identifying which combination of hyperparameters will optimize a model's performance is known as hyperparameter tuning. Based on the prior model comparison, the Random Forest model was determined to be the most suitable for our dataset. To increase its accuracy even more, a grid search approach is applied. With this method, a grid of potential hyperparameter values is created, and each iteration involves investigating a different combination. The model is trained and evaluated using each combination of hyperparameters, and the procedure logs the model's performance. Ultimately, the grid search produces the best model with the most effective hyperparameters, which allows us to fine-tune the Random Forest to achieve accuracy levels closer to the factory values.

Application of GridSearchCV for the Random Forest model's hyperparameter tuning. The hyperparameter values are defined in a parameter grid with multiple combinations, and a grid search is run to find the best set. The optimal parameters are found by fitting the model with the given hyperparameters. This shows that the model performs better when'max_depth' is set to 90,'max_features' to 3,'min_samples_leaf' to 3,'min_samples_split' to 10, and 'n_estimators' to 300. After hyperparameter tuning, the Random Forest model's accuracy increased dramatically from 0.89 to 0.96, indicating that it can now more effectively generate optimal solutions for predicting passenger satisfaction given the available features.

UN-SUPERVISED:

A. Principal Component Analysis

Principal Component Analysis (PCA), a dimensionality reduction technique, to convert the features of an airline dataset into a lower-dimensional representation with the goal of keeping important information while minimizing complexity. PCA is applied using the custom function `pca_transformation}, which yields transformed datasets (`airline_train_pca} and `X_airline_test}). The training set is subdivided into target variable ({Y_airline_train}) and features ({X_airline_train}).

B. K-MEANS

The Method uses the clustering algorithm K-Means to group airline passengers according to their features. The ideal number of clusters is ascertained using the Elbow Method, and in this instance, it is discovered to be four. Next, the training set ({X_airline_train}) is fitted with the K-Means model, and the predicted clusters are applied to the testing set ({X_airline_test}). The accuracy and ROC Area under Curve (AUC) scores that follow show how well the clustering model works. A joint plot of the dataset shows how passengers are distributed among clusters according to flight distance and age. Furthermore, the average values of the features in every cluster are shown, providing information about the traits of travelers in various clusters. The scatterplot illustrates the distribution of passengers in the reduced-dimensional space (PC1 and PC2) based on their assigned K-Means clusters.

C. DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm from scikit-learn to cluster data points in the training dataset (X_airline_train). The DBSCAN model is instantiated with a specified epsilon value (eps=1.1). After fitting the model to the data. There may be a disparity, though, since it runs counter to the explanation's assertion. To comprehend the properties of the generated clusters, more analysis is done on the DataFrame, including grouping and statistical summaries. In addition, the data points are visualized in the space described by principal components PC1, PC2, and PC3 using a 3D scatter plot created with Plotly Express. Colors are used to indicate the "Dissatisfied" variable. It's important to note that the print statement's reported number of clusters differs from the following statement's mention of examining several clusters, pointing to a possible inconsistency in the code or data interpretation.

IV. RESULTS AND DISCUSSION

Logistic Regression-

The Logistic Regression model took about 0.17 seconds to complete and produced an accuracy of 81.22% and a ROC Area under the Curve of 81.91%. The accuracy metric shows how well the model predicts the provided dataset by showing the percentage of accurately predicted outcomes. The efficacy of the model is further demonstrated by the ROC Area under the Curve, which gauges its capacity for class discrimination. The comparatively short training and evaluation times point to computational efficiency. Together, these findings support the Logistic Regression model's capacity to generate precise predictions using the available data, indicating that it is a reasonable option for classification tasks in this situation.

Naive Bayes-

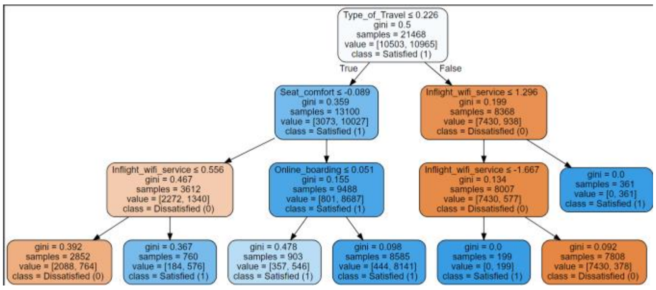
Naive Bayes classifier produced an accuracy of 83.35%. The recorded value of the ROC Area under the Curve is 83.49%., a statistic that evaluates the classifier's capacity to distinguish between classes. The model's computational efficiency is demonstrated by the short training and prediction times, which come in at around 0.03 seconds. Together, these performance measures show how well the Naive Bayes algorithm performs in the specified classification job, balancing computing efficiency and accuracy.

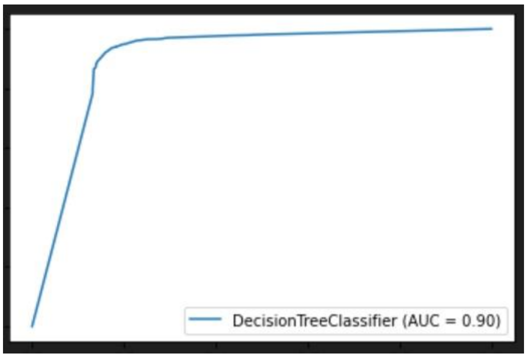
K-Nearest Neighbors (KNN)-

Applying the K-Nearest Neighbors (KNN) method to the dataset produced encouraging outcomes. The model's claimed accuracy of 88.37% indicates that it can accurately predict class labels. The model performs exceptionally well, as evidenced by the ROC Area under the Curve, a statistic that evaluates the model's capacity to distinguish between classes, which is reported at 88.33%. The algorithm's execution time is measured, and the results show that the training and evaluation phases take about 5.86 seconds. Together, these measures demonstrate how well the KNN model performs in identifying underlying patterns in the data and producing precise predictions while requiring a very small amount of processing time.

Decision tree model-

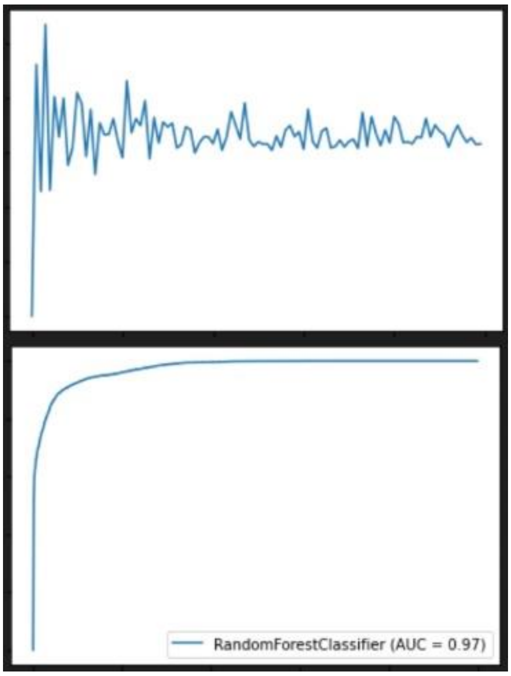
The decision tree model produced a ROC Area under the Curve of 88.11% and an accuracy of 87.43%. The entire training and assessment process took about 0.054 seconds. The ROC Area under the Curve offers information about the model's capacity to distinguish between classes, while the accuracy measure shows the percentage of cases that are properly classified. The model's ability to identify patterns in the data is supported by its comparatively high accuracy. The ability to discriminate is further shown by the ROC Area under the Curve. The computational efficiency of the model is indicated by its quick execution time. Overall, these findings demonstrate the decision tree's excellent performance on the provided dataset in terms of accuracy, discriminative power, and computing economy.





Random Forest-

The Random Forest technique was applied to the dataset, with notable results. With an accuracy of 88.37%, the model demonstrated its ability to accurately categorize cases. With a score of 0.8911, the ROC Area under the Curve (AUC) indicates the model's strong performance and its capacity to distinguish between positive and negative classifications. The efficiency of the algorithm is demonstrated by the fact that the training and evaluation processes took about 2.02 seconds. All things considered, the Random Forest model demonstrates robust prediction skills, yielding excellent discriminative power and accuracy, which makes it a good option for classification tasks on the provided dataset.

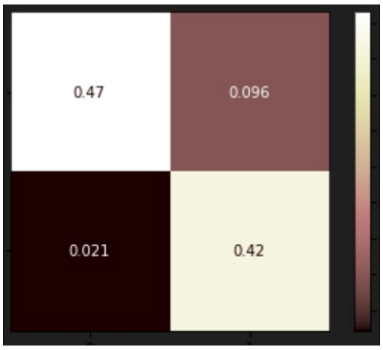


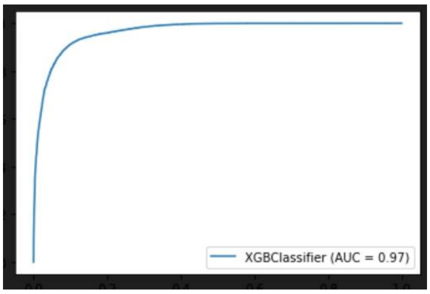
Neural network-

The neural network model's accuracy of 86.12% on the dataset demonstrated its capacity to accurately predict the target variable. With a ROC Area under the Curve (AUC) of 0.87, it appears that sensitivity and specificity are well-balanced. In an efficient evaluation process, the model's performance was assessed in roughly 13.71 seconds. These findings demonstrate the neural network's efficiency in identifying intricate patterns in the data, as shown by its comparatively high accuracy and AUC.

XGBoost-

After applying the XGBoost algorithm on the dataset, encouraging outcomes were obtained. With an accuracy of 88.31%, the model demonstrated its ability to accurately categorize cases. The model's strong performance was further confirmed by the Receiver Operating Characteristic (ROC) Area under the Curve, which is another important parameter for binary classification models. This metric reached 0.8906. The model's computational efficiency is remarkable because it took only about 29.49 seconds to finish the training and evaluation process. Together, these findings demonstrate how well XGBoost performs in terms of producing precise predictions while preserving computing efficiency, which makes it an appealing option for classification tasks.





Adaptive gradient boosting-

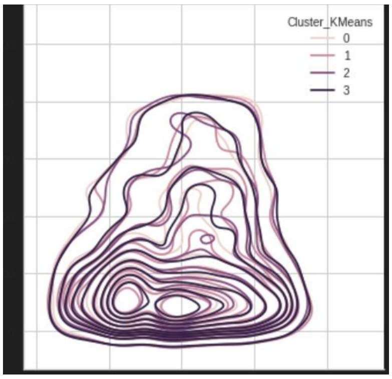
The adaptive gradient boosting model performed admirably on the given dataset, with an accuracy of 88.94%.. At 89.47%, the ROC Area under the Curve—a measure of the model's class discrimination efficacy—was exceptionally high. The training and assessment procedure took 10.04 seconds, demonstrating a respectable level of computing efficiency. In this case, adaptive gradient boosting—a machine learning approach that combines weak learners in a sequential fashion to increase predicted accuracy—worked well. These findings highlight the model's capacity to find patterns in the data, which makes it a good option for classification jobs.

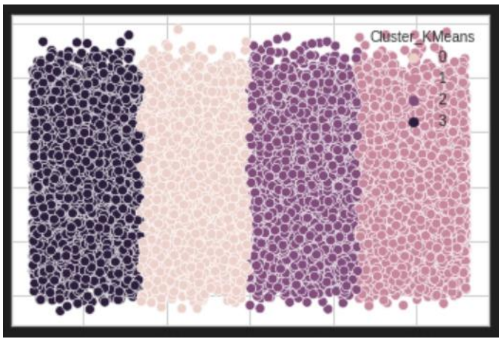
Support Vector Machine-

The Support Vector Machine (SVM) model performed wonderfully on the given data, achieving an accuracy of 88.45%. With a solid ability to distinguish between classes, the Receiver Operating Characteristic (ROC) Area under the Curve, a measure of classification quality, reached 0.8917. A total of 61.32 seconds were needed for training and evaluation, demonstrating the model's computational efficiency. All things considered, these findings demonstrate how well the SVM performs in classification applications, offering great ROC curve performance and high accuracy in a manageable amount of training time.

KMeans-

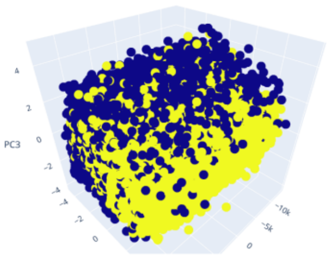
The KMeans clustering algorithm yielded findings with an accuracy of 25.19%, indicating low predictive effectiveness. There is no discernible increase in categorization over random chance, as indicated by the ROC Area under the Curve of 0.50. Four clusters were identified by the program from the data, indicating possible subgroups or patterns. With the selected KMeans clustering configuration, there may be difficulties in accurately capturing the dataset's underlying structure, as indicated by the moderate accuracy and ROC AUC values. It might take more investigation and fine-tuning of the clustering parameters to improve the algorithm's capacity to identify significant patterns in the data.





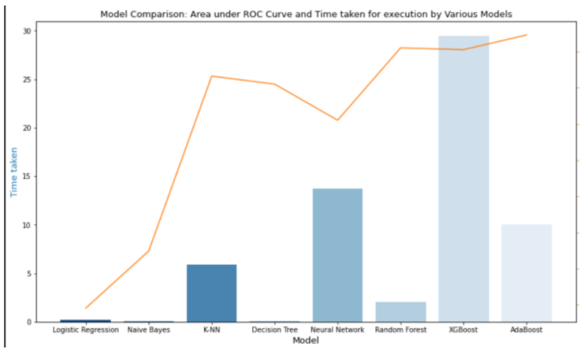
DBSCAN-

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method, which produced the following output: "Number of clusters: 1." An unsupervised machine learning approach called DBSCAN is well-known for its capacity to efficiently handle noise and detect clusters based on the density of data points in the feature space. The single-cluster result implies that the data was viewed by the algorithm as having a single, coherent structure. The interpretation of a single cluster, however, can prompt concerns over the dataset's intrinsic properties or the algorithm's sensitivity to the given parameters. To obtain a better understanding of the underlying patterns and structures in the data, more investigation, analysis, and maybe parameter adjustments may be necessary.



V. CONCLUSION

In conclusion, the assessment of diverse machine learning models focused on two key factors: the Receiver Operating Characteristic (ROC) score and the computing time for training. The ROC score gauges a model's binary classification ability. A comprehensive graph was constructed, featuring training time and ROC score for each model, facilitating a visual and comparative performance analysis. Notably, Random Forest stood out with a high ROC_AUC score (90%) and a shorter training time. Subsequent hyperparameter tuning enhanced the accuracy score to 96%, solidifying Random Forest as the preferred model for its optimal performance.



References

[1]Tsafarakis, S., Kokotas, T. and Pantouvakis, A., 2018. A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. Journal of air transport management, 68, pp.61-75.

[2]Hayadi, B.H., Kim, J.M., Hulliyah, K. and Sukmana, H.T., 2021. Predicting Airline Passenger Satisfaction with Classification Algorithms. International Journal of Informatics and Information Systems, 4(1), pp.82-94.

[3]Archana, R. and Subha, M.V., 2012. A study on service quality and passenger satisfaction on Indian airlines. International Journal of Multidisciplinary Research, 2(2), pp.50-63.

[4]Tahanisaz, S., 2020. Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry. Journal of Air Transport Management, 83, p.101764.

[5]Nurdina, A. and Puspita, A.B.I., 2023. Naive Bayes and KNN for Airline Passenger Satisfaction Classification: Comparative Analysis. Journal of Information System Exploration and Research, 1(2).

[6]Sezgen, E., Mason, K.J. and Mayer, R., 2019. Voice of airline passenger: A text mining approach to understand customer satisfaction. Journal of Air Transport Management, 77, pp.65-74.

[7]Bogicevic, V., Yang, W., Bilgihan, A. and Bujisic, M., 2013. Airport service quality drivers of passenger satisfaction. Tourism Review, 68(4), pp.3-18

[8]Shah, F.T., Syed, Z., Imam, A. and Raza, A., 2020. The impact of airline service quality on passengers' behavioral intentions using passenger satisfaction as a mediator. Journal of Air Transport Management, 85, p.101815.

[9]Knoch, S., Staudt, P., Puzzolante, B. and Maggi, A., 2020, June. A smart digital platform for airport services improving passenger satisfaction. In 2020 IEEE 22nd Conference on Business Informatics (CBI) (Vol. 1, pp. 250-259). IEEE.

[10]Jiang, X., Zhang, Y., Li, Y. and Zhang, B., 2022. Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model. Scientific Reports, 12(1), p.11174.

[11]Archana, R. and Subha, M.V., 2012. A study on service quality and passenger satisfaction on Indian airlines. International Journal of Multidisciplinary Research, 2(2), pp.50-63.

[12]Laws, E., 2005. Managing passenger satisfaction: some quality issues in airline meal service. Journal of Quality Assurance in Hospitality & Tourism, 6(1-2), pp.89-113.

[13]Hong, A.C.Y., KHAW, K.W., CHEW, X. and YEONG, W.C., 2023. Prediction of US airline passenger satisfaction using machine learning algorithms. Data Analytics and Applied Mathematics (DAAM), pp.8-24.

[14]Ban, H.J. and Kim, H.S., 2019. Understanding customer experience and satisfaction through airline passengers' online review. Sustainability, 11(15), p.4066.

[15]Tao, K.N. and Chen, P.J., 2017. Airline passenger safety or customer satisfaction?. International Journal of Business and Economics, 16(2), pp.171-175.