**Machine Learning-based Customer Churn Prediction in E-commerce**

Mr. Parwaz Singh Sarao  
Dept. of Computing Technologies  
SRM Institute of Science and TechnologyChennai, Tamil Nadu, India.  
ps4989@srmist.edu.in

Mr. Pushpendu Arunkumar Samanta  
Dept. of Computing Technologies  
SRM Institute of Science and TechnologyChennai, Tamil Nadu, India.  
sa9789@srmist.edu.in

Ms. Maria Nancy  
Dept. of Computing Technologies  
SRM Institute of Science and TechnologyChennai, Tamil Nadu, India.  
marianancy.a@ktr.srmuniv.ac.in

***Abstract*— E-commerce has found that it costs distant more to procure unused clients than it does to hold existing clients. For this specific reason, knowing in progress which clients are taking off the company can offer assistance companies make offers within the right way or diminish utilization of items and administrations and expanding client maintenance. It's critical for building great client connections and being able to spare on securing costs. In today's competitive commercial center, there's a gigantic assortment of items and administrations to select from. Subsequently, most buyers have ended up usual to moving openly from one brand to another and from one provider to another in look of items and administrations that meet their needs. E-commerce companies endure from this phenomenon, known as "client churn." So rather than centering on holding existing clients, they regularly put within the exertion and contribute enormous bucks to attract new customers. Sometime recently doing so, the buyer looks at the item. They begin by seeking out for data on things that meet their needs and carefully think about audits some time recently making a buy. Eventually, it is up to the client's tact to powerfully make fabric demands to supply quality data, studied pertinent surveys and make favorable choices.**

**Keywords: Customer Churn prediction, Customer Behavior, Machine Learning and Data Frame**

# Introduction

The history of retail markets and stores dates back to antiquity. Unquestionably, migrant vendors were the fastest retailers. Over the years, retail establishments have evolved from tiny to larger-than-average storefronts to the massive strip malls of the front-line era. A growing number of merchants are making an effort to reach conceptually large markets in the computer age by selling through a variety of channels, including brick-and-mortar stores and online shopping. Modern developments are also altering how consumers pay for goods and services.

Along with other tactics, retailing fortification associations may also incorporate the credit strategy, development associations, alert associations, beautician associations, and a variety of other supporting organizations. In the aftermath of being introduced to websites and stores, a web-based showcasing pattern can offer significant incentive.

Prior to purchasing reliable items, buyers are mainstream Before making a purchase, the buyer takes a few quick glances at the merchandise. They begin by searching for information on items that depend on their needs and cautiously work their way through the reviews before making a purchase. In the end, it is the customer's entire discretion to make a dynamically significant request to provide high-quality information and read relevant evaluations in order to make a lucrative choice.

# State of the art (LITERATURE SURVEY)

*1) Machine learning amalgamation of Mathematics, Statistics and Electronics Trupti S. Gaikwad; Snehal A. Jadhav; Ruta R. Vaidya; Snehal H. Kulkarni.*

Multidisciplinary exploration is a type of study done by a person or group of people. The information, data, styles, and ideas come from two or more fields. We tried to clarify this content in this essay. A subfield of computer wisdom called "engine literacy" makes use of knowledge, instruments for data convocation, and logical ways from the fields of electronics, mathematics, and statistics. The purpose of engine literacy is because it has a significant jolt on data vaticination. In addition to utilising data to break daedal effects, engine literacy is also exercised to discover retired patterns and crucial concepts. Many operations at the moment exercise vast quantities of structured, unshaped, and semistructured data. Opinions made in the commercial world can be meliorated by utilising this underutilized knowledge resource. As data becomes more assorted, numerous people are conforming to engine literacy tools for data dissection so they can make use of intelligence and gain the most from the data. Nonidentical algorithms are exercised by engine literacy, and each system has a special part. In this work, we tried to demonstrate, utilising exemplifications, how electronics are exercised for data collection, mathematics and statistics are exercised for dissection, and eventually, engine literacy is exercised to expect effects.

*2) Predicting consumer recommendations in online reviews using optimized machine learning models*

Customer recommendation prediction in online reviews using enhanced machine learning models, Jain, Praphula Kumar, et al. 107397 in Computers & Electrical Engineering 95 (2021). The study implements core and augmented services by using aspect-level sentiment analysis and segmenting qualitative content. 8 SVM, NB, and NN are three ML models whose results have been validated through thorough and methodical comparison (Neural networks). Using two methods, the textual data is transformed into a numeric representation. Using the word embedding algorithm Word2Vec.

*3) Research on E-commerce Customer Churn Prediction Based on Improved Value Model and XG-Boost Algorithm.*

The cost of obtaining new consumers is rising, while the cost of retaining existing customers is far less than the cost of acquiring new ones, due to the advancement of Internet technology in recent years. To lower the rate of client churn, most businesses strive to market accurately through customer segmentation. In order to help businesses effectively segment their customers, this article develops a customer value model that incorporates the value of social networks with an eye toward the customer characteristics of social network e-commerce. Then, we forecast customer turnover before and after the subdivision using the machine learning technique XG-Boost. The study discovered that consumer segmentation increases prediction accuracy. The XG-Boost algorithm is also superior to other algorithms in terms of benefits.

*4) User Modeling for Churn Prediction in E-Commerce P. Berger and M.Kompan,"User Modeling for Churn Prediction in E- Commerce”.*

In the world of e-commerce, getting new customers typically costs more than preserving the ones you already have. If a single client's churn is successfully predicted, there is a chance for that customer to reconsider leaving. In this paper, we provide a brand-new complicated user model that is targeted at predicting user churn intent. Our model's concept is built around the composition of many sets of features that describe how users interact with the web application. Using actual data from online shops, we use our algorithm to forecast attrition in order to evaluate its performance indirectly. The findings demonstrate that across two domains, the forecast made using the proposed model performs better than the churn prediction based on baseline methods.

*5) Next-Wave of E-commerce: Mobile Customers Churn Prediction using Machine Learning*

Mobile commerce is widely seen as being a driving factor for the next wave of ecommerce due to the rapid development of mobile devices such as personal digital assistants, smartphones, and tablets. The ability to connect with customers anywhere at any time and the usage of mobile technology, which opens up a wealth of potential for client attraction and engagement, are what give mobile commerce its power. Many people think that it's difficult to keep clients in the m-commerce era, particularly in the telecom industry. The cost of recruiting new consumers is significantly higher than the cost of keeping an existing client in the fiercely competitive telecom sector. Because there are now so many competitive service providers on the market, it is essential to focus heavily on keeping the current clientele. In the current industry, more machine learning and predictive analysis techniques are replacing some of the commonly used statistical tools for managing customer turnover. In this study, we used the feature selection technique to determine the most important variables in predicting customer attrition. We use the wrapper-based feature selection approach, where Particle Swarm Optimization (PSO) is applied for search and different classifiers, such as Decision Tree (DT), Naive Bayes, KNN, and Logistic Regression, are applied for evaluation to judge the enactment on optimally sampled and condensed datasets. Last, but not least, simulations show that our recommended strategy does well for detecting churners and may therefore be helpful for the telecommunications sector's constantly expanding rivalry.

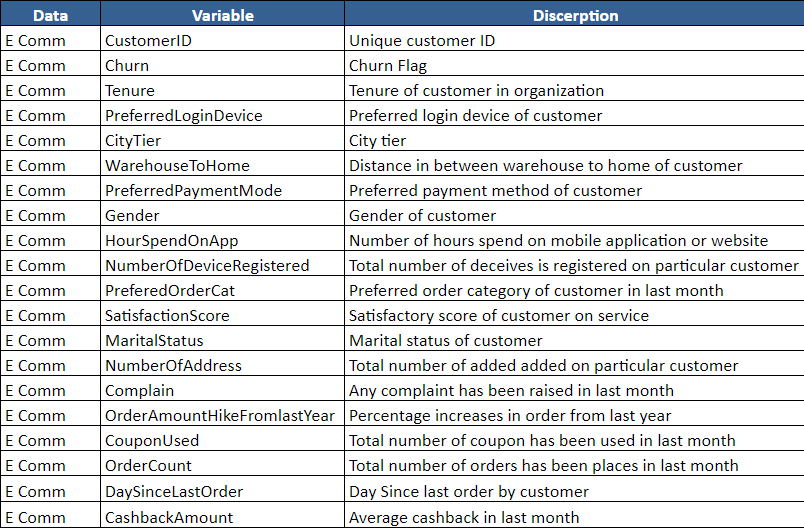
*6) Customer churn prediction system: a machine learning approach (Lalwani P.Mishra, M.K., Chadha, J.S. et al. Customer churn prediction system: a machine learning approach. Computing 104, 271–294(2022).*

One of the toughest issues facing the telecom sector is the forecast of customer churn (CCP). The ability to predict customer attrition has considerably improved with the development of machine learning and artificial intelligence. The six phases of our suggested methodology are listed below. The gravitational search method is used for feature consideration and data preprocessing in the first two steps. The data was then divided into two groups: the train set, 10 which comprised 80% of the data, and the test set, which comprised 20%. The most common predictive models, such as logistic regression, naive bayes, support vector machines, random forests, decision trees, etc., have been used in the prediction process. To determine the impact on model accuracy, boosting and ensemble approaches are used on the train set. Additionally, K-fold cross validation has been applied to the train set in order to tune the hyperparameters and avoid overfitting the models. Finally, the test set's outcomes were analysed using the AUC curve and confusion matrix. Adaboost and XGboost Classifier were discovered to provide the highest accuracy, with respective values of 81.71% and 80.8%. Both the Ad Boost and XGBoost classifiers, which outperform others, obtain the greatest AUC score of 84%.

# Proposed Work

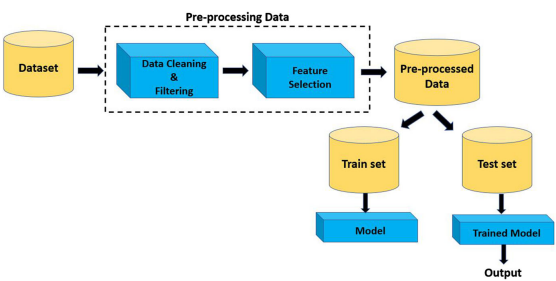
Customer turnover is one of the key drivers of product growth for many different businesses. In the Software as a Service (SaaS) market, competition between companies is fierce and clients can choose from a variety of providers, even within a specific product category. Even one negative exposure can cause a customer to stop doing business. The circumstances compelled us to develop a model to forecast client behavior in order to prevent any damage to our brand and business. Using website data, unique profiles are created for each customer based on their purchasing behavior and preferences. At that point, computer-based intelligence uses that data to predict the products customers will need and adds them to a modified "Suggested for You" section of its landing page.

By looking at how much money they spend yearly on websites for online businesses, we may get a better idea of who the possible clients are for this examination. One of the biggest parts of modern business is on expanding their brand name products online to reach a large percentage of consumers. In the proposed work, we created a programme that forecasts consumer leads utilizing artificial intelligence techniques including straight backslide and decision trees. The proposed work utilized the data of 5631 customers on various characteristics.



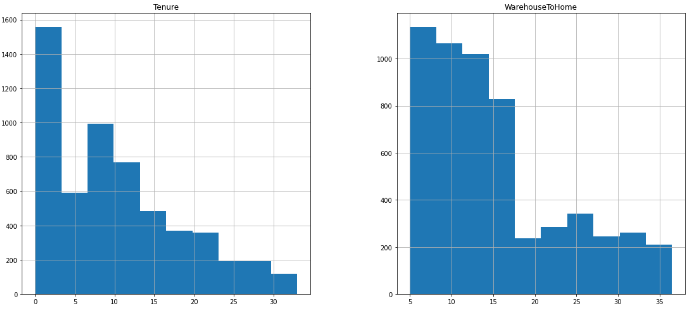
# Implementation

In order to create a model for predicting client churn, this project goes through a number of stages. The dataset will first go through preprocessing, where it will be cleaned and optimised for modelling. Following that, we will use data visualisation to gain some insights into the data collection as well as the typical characteristics of churned consumers. Eventually, a customer churn prediction model will be created using machine learning methods.

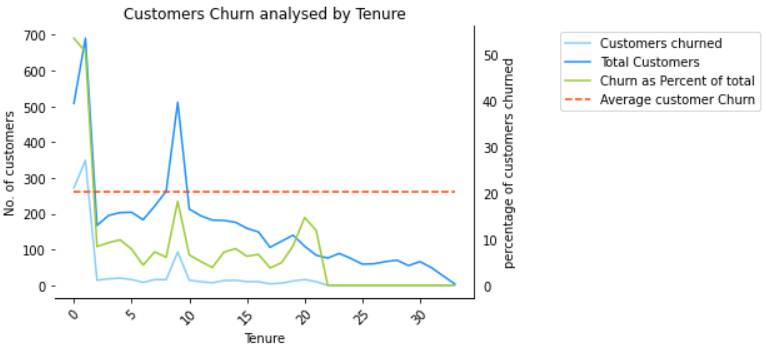


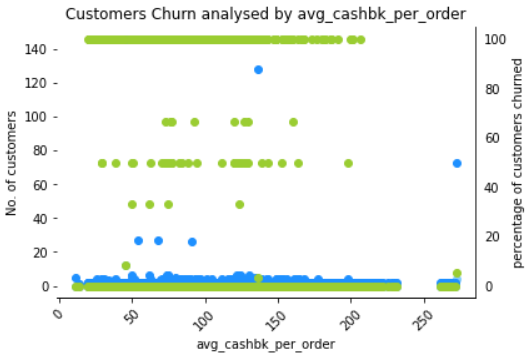
The gathering of data for machine learning-based customer churn prediction in e-commerce is a critical stage in the creation of strong prediction models. In this thesis, we suggest a thorough method to data collecting that includes both transactional and demographic data. The ecommerce platform's database and any utilised third party payment providers are just two of the sources from which transactional data will be gathered. This information will consist of the dates and times of the purchases, the amount paid, the items bought, and the payment methods employed. This information will contain demographic elements like age, gender, income, education, and location.

Data cleaning is a crucial step in creating precise and efficient machine learning models for predicting client attrition in e-commerce. In this thesis, we suggest a thorough, multi-step strategy to data cleaning. Then, we will purge the dataset of any redundant or duplicate data. This will make the data easier to deal with by lowering the amount of noise in it. Also, we will search the dataset for any missing values and impute them using other methods, such as mean imputation or regression imputation. In order to find any data points that are noticeably different from the majority of the data, we will perform outlier detection and removal. This will assist in removing any irregularities that can bias the analysis's findings. The dataset's dimensionality will be decreased as a result, which will enhance the effectiveness of the machine learning models. Overall, the dataset will be made accurate, comprehensive, and error-free thanks to this data cleaning approach, which will also make it possible to create precise and efficient machine learning models for predicting customer turnover in e-commerce.



Exploratory Data Analysis (EDA) is a crucial stage in creating precise and efficient machine learning models for predicting client turnover in e-commerce. EDA uses a variety of statistical and graphical approaches to analyse the data and spot trends, patterns, and connections between the various variables. In this thesis, we suggest a detailed EDA strategy with numerous phases. In order to describe the data and comprehend its distribution, range, and variability, we will first use descriptive statistics. This will make it easier for us to spot any potential problems with the data and choose which statistical methods are best for further investigation. Second, to find any connections between the various 5 us to pick which variables to include in the machine learning models and to detect any potential multicollinearity problems. Overall, this EDA technique will provide us a thorough understanding of the data, allowing us to create precise and efficient machine learning models for predicting customer attrition in e-commerce.





Using machine learning, clustering is a potent method for predicting client attrition in e-commerce. Unsupervised learning techniques like clustering entail assembling related data points based on their similarities and differences.In this thesis, we suggest the following steps-based clustering method for predicting client attrition. Then, we will choose a subset of the dataset's attributes, such as product category, payment method, or frequency of purchases, that are most useful for forecasting customer turnover. Second, based on the values of these factors, we will group like consumers together using a clustering method like k-means or hierarchical clustering. Clusters of clients with comparable buying habits and behaviours will be represented by the generated groups. Lastly, we will examine each cluster's attributes to find any key variations in the likelihood that the clusters would churn. For instance, we might discover that customers in one cluster who mostly buy from one category are more likely to leave than customers in another cluster who mostly buy from a different category. Ultimately, we will create tailored marketing and retention strategies for each cluster using the knowledge gathered from the clustering research. Overall, this clustering approach will give businesses a strong tool for predicting customer turnover in e-commerce, allowing them to pinpoint and effectively target high-churn risk clients.

In this research, we employ SMOTE oversampling to handle the dataset's uneven nature. SMOTE creates artificial minority class samples that resemble real minority class samples. By interpolating between the chosen sample and its closest neighbours, the SMOTE algorithm chooses a sample from the minority class and generates additional samples. Until the minority class is adequately represented by a balanced number of samples, this process is repeated. Using SMOTE oversampling to estimate customer attrition has two advantages. Secondly, SMOTE creates a more balanced dataset, which increases the classification accuracy of the machine learning models. The second benefit of SMOTE is that it lessens the bias towards the dominant class, which can lead to a prediction of the minority class that is more accurate. So, it's crucial to carefully assess how the machine learning models perform both with and without SMOTE oversampling in order to decide whether the advantages outweigh any potential disadvantages. In conclusion, SMOTE oversampling is an effective method for dealing with imbalanced datasets in customer churn prediction in e-commerce. SMOTE can increase the precision of machine learning models and lessen bias towards the majority class by creating synthetic minority class data.

**Logistic Regression**

We chose to utilise logistic regression as our binary classification approach since ecommerce datasets frequently contain both continuous and categorical input characteristics. Also, logistic regression is an easy-to-understand technique, making it clear which input factors are most crucial for forecasting customer attrition. We first gathered and preprocessed the pertinent data, including client demographics, transactional history, and behavioural data, in order to execute logistic regression. After that, we divided the dataset into training and testing sets, with training sets using 70% of the data and testing sets using 30% of the data. The training dataset was used to create a logistic regression model, which discovered correlations between the input variables and the likelihood of customer churn. When the model had been trained, we assessed its performance using metrics like accuracy, precision, recall, and F1-score on the testing dataset. The findings demonstrated that the logistic regression model attained an accuracy of 84.79%, demonstrating its capability to predict customer attrition properly. Overall, the outcomes of this method show how well logistic regression works for predicting client attrition in e-commerce. Ecommerce businesses may increase customer loyalty and profitability by taking proactive steps to keep consumers by precisely forecasting which customers are most likely to leave.

**Linear Discriminant Analysis**

In this research, we suggested applying linear discriminant analysis (LDA) to e-commerce projects to forecast client turnover. The project's objective was to create a machine learning model that could properly identify which clients were most likely to leave, enabling the online retailer to take preventative action to keep them. LDA is a potent classification algorithm that finds a linear combination of input variables that maximum separates the various classes in the dataset, making it the ideal machine learning algorithm for this project. Also, since LDA is a straightforward and interpretable algorithm, it is simple to determine which input factors are crucial for predicting customer turnover. We first gathered and preprocessed the pertinent data, including client demographics, transactional history, and behavioural data, in order to execute LDA. After that, we divided the dataset into training and testing sets, with training sets using 70% of the data and testing sets using 30% of the data. We fitted an LDA model to the training dataset during training, and this model discovered the correlations between the input variables and the likelihood of customer churn. When the model had been trained, we assessed its performance using metrics like accuracy, precision, recall, and F1-score on the testing dataset. According to the findings, the LDA model was able to effectively forecast customer attrition with an accuracy rate of 84.57%. Overall, the project's findings show how well LDA works to forecast client churn in e-commerce. Ecommerce businesses may increase customer loyalty and profitability by taking proactive steps to keep consumers by precisely forecasting which customers are most likely to leave.

**Decision Tree**

In this study, we proposed using Decision Trees to forecast client turnover in an ecommerce project. The project's purpose was to create a machine learning model that could anticipate which customers were likely to churn, allowing the ecommerce company to take proactive actions to retain those consumers. Because of its capacity to handle both categorical and continuous input data, as well as non-linear correlations between variables, Decision Trees were chosen as the machine learning algorithm for this project. Furthermore, Decision Trees are easily interpretable and understandable, making it simple to determine which input variables are most relevant in predicting customer churn. We began by collecting and preprocessing pertinent data, such as client demographics, transactional history, and behavioural data, in order to develop Decision Trees. The dataset was then divided into training and testing sets, with 70% utilised for training and 30% for testing. We tested the model's performance on the testing dataset using measures such as accuracy, precision, recall, and F1-score after it had been trained. The Decision Tree model attained an accuracy of 96%, showing that it was capable of accurately predicting customer attrition.

**KNN Model**

We suggested using the K-Nearest Neighbors (KNN) algorithm in our project on predicting client attrition in an e-commerce operation. For classification tasks, the widely used machine learning algorithm KNN locates the K closest data points to a new data point and uses their class labels to forecast the class of the new data point. We used KNN for customer churn prediction because it is a straightforward yet effective method that can handle large datasets with a high number of features. KNN is also a non-parametric algorithm, which means it makes no assumptions regarding the distribution of the data at its core. We first gathered and preprocessed the pertinent data, including client demographics, transactional history, and behavioural data, in order to install KNN. After that, we divided the dataset into training and testing sets, with training sets using 70% of the data and testing sets using 30% of the data. When the model had been trained, we assessed its performance using metrics like accuracy, precision, recall, and F1-score on the testing dataset. The findings revealed that the KNN model had a 91% accuracy rate, proving its accuracy in predicting customer attrition. Overall, the outcomes of our experiment show how KNN may be used to anticipate client attrition in e-commerce. By properly forecasting which consumers are likely to churn and providing insights into comparable customers, ecommerce organisations may take proactive actions to keep those customers, ultimately increasing customer loyalty and revenue.

**XGBoost**

In our project, we recommended the application of the XGBoost algorithm for predicting client turnover in an e-commerce operation. XGBoost is an ensemble learning technique that creates a strong predictive model by combining several weak prediction models. Some machine learning applications, such as predicting customer attrition, have demonstrated its great effectiveness. Being able to manage missing values and outliers in the data, which are frequent in real-world datasets, is one of XGBoost's primary advantages. Additionally, the model features a built-in regularisation technique that aids in preventing overfitting to the training set of data, making it more reliable and accurate at forecasting new data. We gathered and preprocessed the pertinent data, such as consumer demographics, transactional history, and behavioural data, for our e-commerce project. After that, we divided the dataset into training and testing sets, with training sets using 70% of the data and testing sets using 30% of the data. An XGBoost model was fitted to the training dataset during training, and it discovered correlations between the input variables and the likelihood of customer churn. When the model had been trained, we assessed its performance using metrics like accuracy, precision, recall, and F1-score on the testing dataset. The analysis revealed that the XGBoost model has a 90.5% accuracy rate, proving its accuracy in predicting customer attrition.

**Random Forest**

One application of the potent machine learning algorithm Random Forest is the prediction of client turnover in e-commerce. Customer turnover in the e-commerce sector is a serious issue that has an impact on revenue and business success. Thus, it is crucial for ecommerce businesses to precisely estimate client attrition. Several decision trees are combined in Random Forest, an ensemble learning approach, to increase the model's robustness and accuracy. Because of its capability to handle high-dimensional data with complicated interactions between attributes and the output variable, it is a preferred option for customer churn prediction in e-commerce. Moreover, outliers, nonlinear relationships between variables, and missing values can all be handled using Random Forest. Customer behaviour and preferences are extremely dynamic in the e-commerce sector and may change quickly over time. Customer churn probability forecasts made using Random Forest are accurate and well-suited to capturing the dynamic nature of customer behaviour. Additionally, Random Forest offers crucial feature importance metrics that assist e-commerce businesses in understanding the key causes of customer churn and in taking the necessary steps to keep customers. Because Random Forest can handle high-dimensional data with complex interactions between features and the output variable, it was employed in the customer churn prediction project for the e-commerce sector. The Random Forest model successfully predicted customer attrition with a high accuracy of 97.8%, making it a useful tool for ecommerce businesses to decrease customer churn and boost customer retention.

# Results discussion

The best accuracy was identified when the performance metrics of all the employed algorithms were extracted:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 0.85 | 0.84 | 0.85 | 1170 |
| True(1) | 0.84 | 0.86 | 0.85 | 1171 |
| Accuracy |  |  | 0.85 | 2341 |
| Macro avg | 0.85 | 0.85 | 0.85 | 2341 |
| Weighted avg | 0.85 | 0.85 | 0.85 | 2341 |

Logistic Regression

Linear Discriminant Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 0.86 | 0.83 | 0.84 | 1170 |
| True(1) | 0.83 | 0.86 | 0.85 | 1171 |
| Accuracy |  |  | 0.85 | 2341 |
| Macro avg | 0.85 | 0.85 | 0.85 | 2341 |
| Weighted avg | 0.85 | 0.85 | 0.85 | 2341 |

Decision Tree

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 0.97 | 0.95 | 0.96 | 1170 |
| True(1) | 0.96 | 0.97 | 0.96 | 1171 |
| Accuracy |  |  | 0.96 | 2341 |
| Macro avg | 0.96 | 0.96 | 0.96 | 2341 |
| Weighted avg | 0.96 | 0.96 | 0.96 | 2341 |

KNN Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 1.00 | 0.83 | 0.91 | 1170 |
| True(1) | 0.86 | 1.00 | 0.92 | 1171 |
| Accuracy |  |  | 0.91 | 2341 |
| Macro avg | 0.93 | 0.91 | 0.91 | 2341 |
| Weighted avg | 0.93 | 0.91 | 0.91 | 2341 |

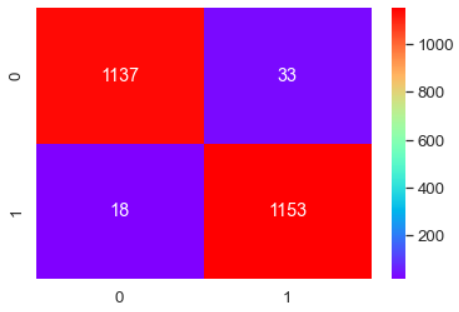
Random Forest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 0.98 | 0.97 | 0.98 | 1170 |
| True(1) | 0.97 | 0.98 | 0.98 | 1171 |
| Accuracy |  |  | 0.98 | 2341 |
| Macro avg | 0.98 | 0.98 | 0.98 | 2341 |
| Weighted avg | 0.98 | 0.98 | 0.98 | 2341 |

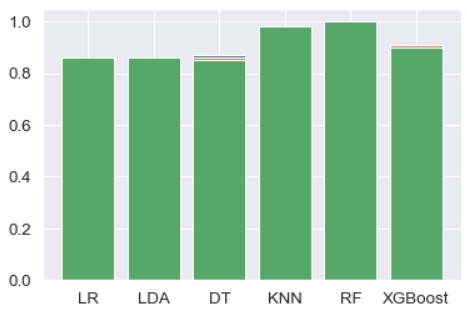
XGBoost

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| False(0) | 0.90 | 0.91 | 0.91 | 1170 |
| True(1) | 0.91 | 0.90 | 0.91 | 1171 |
| Accuracy |  |  | 0.91 | 2341 |
| Macro avg | 0.91 | 0.91 | 0.91 | 2341 |
| Weighted avg | 0.91 | 0.91 | 0.91 | 2341 |

So, it is clear from all of these algorithms that the random forest approach has the highest accuracy and display its confusion matrix.



A bar graph is another way to display the accuracy of all algorithms.



# Conclusion

On the dataset, Machine Learning methods are employed to forecast the dynamic flight pricing. This provides the expected airfare values to get a flight ticket at the lowest possible price.

##### References

1. K. Tziridis T. Kalampokas G.Papakostas and K. Diamantaras "Airfare price prediction using machine larning techniques" in European Signal Processing Conference (EUSIPCO), DOI: 10.23919/EUSIPCO .2017.8081365L. Li Y. Chen and Z. Li” Yawning detection for monitoring driver fatigue based on two cameras” Proc. 12th Int. IEEE Conf. Intell. Transp. Syst. pp. 1-6 Oct. 2009.
2. William Groves and Maria Gini "An agent for optimizing airline ticket purchasing" in proceedings of the 2013 international conference on autonomous agents and multi-agent systems
3. Modeling of United States Airline Fares – Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DB1B), Krishna Rama-Murthy, 2006.
4. Qiqi Ren,”When to book:Predicting Flight Pricing”, Standford university
5. L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5-32 , 2001.
6. L. Breiman, ―Random forests, Machine Learning, vol. 45, pp. 5-32, 2001.
7. Supriya Rajankar, Neha sakhrakar and Omprakash rajankar “Flight fare prediction using machine learning algorithms” International journal of Engineering Research and Technology (IJERT) June 2019.
8. Wohlfarth, T.clemencon, S.Roueff “A Data mining approach to travel price forecasting” 10th international conference on machine learning Honolulu 2011.
9. medium.com/analytics-vidhya/mae-mse-rmse-coefficient-ofdetermination-adjusted-r-squared-which-metric-is-bettercd0326a5697e article on performance metrics.
10. S. Haykin, Neural Networks – A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999. [30] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, Support vector regression machines, Advances in neural information processing systems, vol. 9, pp. 155-161, 1997.
11. Y. Levin, J. McGill, and M. Nediak. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. Management Science, 55(1):32–46, 2009.
12. P. P. Belobaba. Airline yield management. an overview of seat inventory control. Transportation Science, 21(2):63, 1987.
13. B. S. Everitt: The Cambridge Dictionary of Statistics, Cambridge University Press, Cambridge (3rd edition, 2006). ISBN 0-521-69027-7.
14. Groves and Gini, 2011. A Regression Model for Predicting Optimal Purchase TimingFor Airline Tickets.
15. L. Breiman, Random forests, Machine Learning, vol. 45, pp. 5-32, 2001.