Analysis Understanding Online Shoppers' Purchase Intentions: Using Machine Learning

Harshitha Gyada
2203A52025
Department of Computer Science Engineering
SR University
Hanamkonda , India
2203A52025@sru.edu.in

***Abstract*—**This project focuses on predicting online shoppers' purchasing intentions using machine learning. We collect and pre processed data, engineer features, and select the best-performing models. After rigorous testing, we deploy the model to improve e-commerce strategies, increase user satisfaction, and boost sales. Continuous monitoring and proper documentation ensure long term success. The project aims to empower businesses with data driven insights to enhance their online shopping 5 experience and drive growth.

***Index Terms*—Astrophysical randomForestClassifier, supportvectormachines (SVM), kearest Neighbors (KNN), Decision Trees, Logistic Regression, Feature Importance, Data Preprocessing and Scaling, Oversampling Techniques, Model Evaluation Metrics**

INTRODUCTION:

In the course of the endless evolution of digital space, the way consumers consume has changed tremendously from off line to online. It would not be an overstatement to say that online shopping has become a new trend against which some businesses are now standing to survive in the hyper competitive e commerce world. Identifying those key factors that cause online shoppers to be more interested in businesses appealing them is of the utmost importance for these businesses. In this article, we take a trip to the world of the online shopping habitat of the shoppers intentions through the help of machine learning(ML) model that can make data-driven decisions thus helping to craft targeted marketing strategies, improving the user experience and driving online sales. In the era of the digital where all major business decisions makers require data-driven decision-making, ML has proven as the courageous mainstream tool, capable of unfolding the sophistication of caprices and preferences of consumers. Now that has dominantly become everything that is the case in convenience and accessibility this particular type of online shopping has occupied consumer's attention from all across the globe. This project is Nailing down the idea of within E-commerce and seeking to unveil the potency of ML as a means of revealing what underlies online shoppers' purchasing decisions. This knowledge is irreplaceable for businesses in order to stay perking up and succeeding in the industry that gradually goes digital.

Machine learning with its sophisticated algorithms and data analytic powers is ready to be an important player in unraveling the issues of online commerce. Through the use of ML, this project is able to search the hidden factors operating along the online shoppers fluctuation axis. This will give businesses the opportunity to move with the market movements and proliferate by feeding off the information to decide their strategies as well as revise their offers accordingly. The ML is likely to allow researchers to be ahead of the curve in both the case of technological growth and is also at the ready to reveal deeper insights that go beyond traditional research method.

The aim of the study is determining which of online factor (influences) that drive purchase intent of shoppers. The goals encompass the elements like product quality, price, user friendly website, reviews and the social media positivity. Thereby, the application of machine learning algorithms models will not only identify these determinants but will also examine their magnitude enabling companies to distinguish what to prioritize or what should come first. The expected outcomes of this project are twofold: making it possible for businesses to comprehend the key data and to profit from Sales successes and keep the development of consumer behavior in online shopping through their research.

The scene of electronic commerce has shifted away from using the traditional store concepts to redesign the shopper experience—this has no doubt made the internet shopper's purchasing intentions critical now more than ever. In this project, the world of consumer behavior will be demystified as its root lie in the beneficial phenomenon of machine learning. Through richer Machine Learning algorithms as well as implementing the method of large data analysis, it aims to get a deeper understanding of the complex network of factors that affect how online shoppers behave, and it provides businesses with information that can be used as a compass to catch up in the thickly packed digital marketplace.

LITERATURE REVIEW

1.Ali Khandakar, A. Islam, Salekul Islam, and Swakkhar Shatabda presented the customer purchases intention predictions model by using feature selection techniques and oversampling methods with higher performance. The authors ruled superior performance of their approach against prevailing work by Bakti et al.,and Song et al., , TNR and f1-score and many more.

2.Tajba Asgher employed ML algorithms to do online shopping customer intensions. Analyzed the purchase intentions of customers placing a focus on machine learning algorithms. I created a user profile system that allowed the discovery of patterns from customer history and the pageviews of the purchases they made. There were clustering schemes and clickthrough data used for modeling customers. Smeared data mining tools to predict future customer purchase propensity. Please don't waste my time

3.Zhenyu Liu, Xinyi Ma. An original approach to using machine learning for customer purchasing behavior assessment. Offered a hybrid model of prediction using Stack method that combines decision trees in order to make a more informed decision. Combine Light GBM, XGB, and RF with regularization, GBM L1, for precise predictions. Evaluation mock-up and verification with real retail sales data is underway.

4.Thése Second Thousand, a book authored by Susanna Rossum and Raphael Drake. The prediction of online shoppers intention at present by Sakar et al is understandable. The move of the consumer journey was examined through attention based recurrent neural networks by Zhou et al.The use machine learning and deep learning models bythe Zhou et al. to study the buying intention of the online customers. Identified the most significantly contributing factors affecting consumer's purchasing attitude. Apply the Shapley Additive Explanation process to figure out the most importance features. The research will involve the assessment of several ensemble learning models and the comparison of deep learning models performance.

5.Thus, through the online shopping study, Daksh Kapoor, Achirangshu Chakraborty, Sunita Daniel utilized machine learning algorithms to group customers according to a number of factors like whether or not they shopped online regularly on the internet. Carried out the analysis data using ensembling method to recognize intention in shopping. Correlated revenue, customer and shopping visits for online shopping to the given month. Apply Gradient Boosting for more precise prediction of user patterns.

6.Jiangtao Qiu. Propose a COREL model for customer purchase behavior predicting in e-commerce. Investigated the possible causes of shopping decisions in e-commerce and look for the methods to quantify customers requirements, products popularity and their preferences. Developed a tool where the analysis of purchase behavior was displayed in the form of COREL predictive model.

7.Priyank Sirohi, Niraj Singhal, Pradeep Kumar, Mahboob Alam. Proposed a customer behaviour model based on gradient boost approach Gradient boosting approach-based customer behaviour model proposed. Elaborate a mathematical process used for behavior analysis of customers.

8.Shaifali Yadav, an M.Tech scholar of Devi Ahilya Vishwavidyalaya (DAVV), India. Solve the problem of long delivery times with using machine learning algorithms on e-commerce clickstream data. Split data into train and test sets, taking note of Accuracy, Precision, Recall, and f1 score of each algorithm. Explored hyperparameter tuning for the improvement in accurateness. Each of them visualized the data with logisticregression, decisiontree, randomforest, more algorithms.

9.Andrew frazier, fatbardha maloku, xinzi li, yichun chen, YejiJung, BahmanZohuri.Web browsing behaviors analytics was the model developed to predict customer purchases.Thanks to machine learning, one was provided with insight into online shoppers purchasing intents. Modelling process done through Scikit learn Pipeline tool for data preprocessing and also classification. Further examined clasical representation of numerical features with target class through Exploratory Data Analysis.

10. Md Shahriar, Said be-Shayeh, Fhardool Insa:ir, Md Satv be-Pusuri. Offered new model of ML to analyze online behaviors.Developed a new model on ML to predict intention of purchase online. Experimented with functionality both feeding and controlling the data. The outcomes of diverse classifiers for the online forces' purchasing cases were studied and analyzed. Presented online consumers intention of purchase forecast via the proposed model.

11.Yap Chau Tean. Specifically, he looks for the applications of data mining approaches in the e-commerce field. Yap Chau Tean obtained the research focusing online purchasers' purchase intentions' predictability. Performed six supervised machine learning algorithms classification of online customers' data Conducted data pre-processing with over sampling, under sampling, and hybrid sampling; Tested Ada Boost and Bagging but saw no difference in prediction; Sustained unbalanced data set and its implications.

12.Paola Furtado, Rosales M´ aster En Econom´ ıa, Diego Mar´ ın Santos.Applied supervised machine learning to predict online purchase intentions. Applied supervised machine learning techniques to predict purchase intention. Used database with 18 variables divided into training and evaluation sets. Balanced training set by random downsampling.Configured models with two sets of predictors using attribute selection

13.Yi Lim, Abdullah Osman, Shahrul Nizam Salahuddin, Abdul Rahim Romle, SafizalAbdullah.Conducted research on Shopping behavior, purchase intention, and subjective norms.Analyzed data using SPSS and AMOS for hypothesis testing. Developed hypotheses based on previous studies on online shopping behavior. Explored subjective norm and usefulness in shopping behavior.

14.Muhammadadlansyah Muda, radha, Ayu Aswari, Muhammad Ahsan. Authors focus on predicting online purchasing intention using machine learning.Used binary logisticregression, decision tree, and random forest methods. Employed oversampling and feature selection for model refinement. Identified key revenue factors: new visitors, low bounce rate, etc. Achieved 88.21 accuracy with random forest model using cross-validation.

III.PROPOSED APPROACH

1.Data Description

Analyzing e commerce online shoppers' buying intention in machine learning involves concerning what consumers' do online and by this they will be able to improve personalized recommendations and soon more sales. Machine learning can predict Zalando's customer's intention of purchase before the session is ended. The discounts and offers can be tailored made while an anonymous customer browse in the e-shop, which is a crucial for many e-shoppers. With applying the information through digging data mining techniques, the satisfaction, efficiency, and engagement of the customers in the shopping process with high rates of conversion can be increased,and it will be competitive success. The algorithms used are SVM ensembles, decision trees, neural networks and so on to identify consumers attitude towards buying a product online so, retailers can expand their customer base and channels can remove impediments to make their platforms more efficient. This strategy is crucial in order to meeting needs of anonymous shoppers, which results in better quality of the user experience and financial progress of e-commerce platforms.

	Administrative	Administrative_Duration	Informational	Informational_Duration	ProductRelated	ProductRelated_Duration	BounceRate
0	0	0.0	0	0.0	1	0.000000	0.2
1	0	0.0	0	0.0	2	64.000000	[illegible]
2	0	0.0	0	0.0	1	0.000000	0.20
3	0	0.0	0	0.0	2	2.666667	0.05
4	0	0.0	0	0.0	10	627.500000	0.02

2. Data analysis

These distributions reveal how features influence in the decisions making problem. Depending on not overlapping and individual characteristics of the classes, the features will be discriminative for classification . Some specific features could have substantial overlap which, in turn, became of less use for the class distinction.

These histograms can be used to do a thorough feature exploration, model selection, and composing of feature set to see the final results .

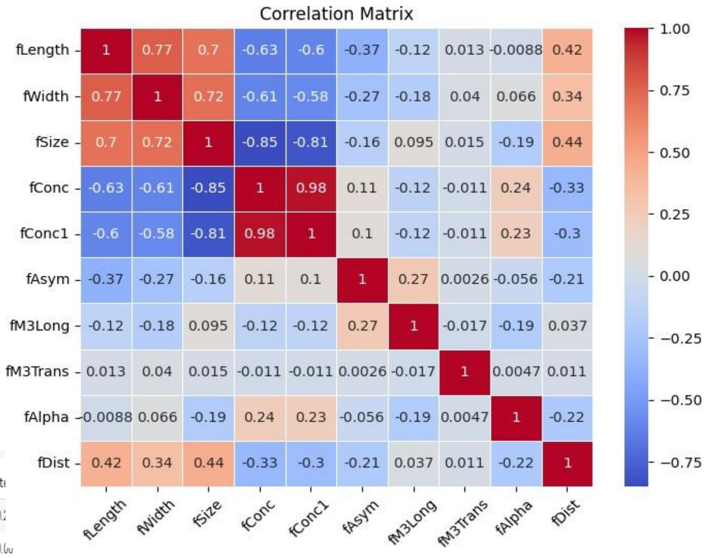
Correlation Matrix

A correlation matrix, also known as a matrix of correlation coefficients, it is a table or matrix that displays Correlation Coefficients amount multiple variables in a data set. Correlation coefficients quantify the degree and direction of the Relationship between pairs variables. In a co-relation matrix, every cell tells a co-relation between the two variables.
Correlation coefficients typically fall in the range -1-1, where:.

1 indicates positive correlation: When 1 variable increases, so does the other also increases proportionally.

-1 indicates negative correlation: When 1 variable increases, so the other decreases proportionally.

0 indicates nolinear correlation: The variables are not related in a linear manner.



3. Data preprocessing

Data cleaning and shaping are essential parts feature extraction and selection steps, which go on structuring raw data in a format acceptable for the model. It is of epic importance for it presents a purpose to provide crucial data and machine learning in order for the model to pick up on the well-defined patterns. Here is a detailed explanation of the various aspects:

Data cleaning:

Handling missing values: Identify and handle missing data, either by deleting rows or filling in missing values using techniques such as Mean, Median, or Interpolation. There are no null values in this dataset.

Data transformation:

Feature Scale: Normalize or standardize numeric features to Ensure different feature are on a same scale. Techniques include min-max scaling and Zscore normalization.

Data distribution:

Training, Validation and Testing Split: Divide the data set into categories of training, validation and testing. The training set is of training the model, the validation set is used to fine-tuning the hyperparameters and the test set is used to performance testing of the model..

Working with string data :

Text Preprocessing- For natural language processing tasks, preprocess text data by tokenizing, removing stop words, deriving or lemmatizing, and converting text to numerical representations (eg, TF-IDF or word embedding).

PageValues	SpecialDay	OperatingSystems	Browser	Region	TrafficType	Weekend	Revenue
0.0	0.0	0.000000	0.000000	0.000	0.000000	0.0	0.0
0.0	0.0	0.142857	0.083333	0.000	0.052632	0.0	0.0
0.0	0.0	0.428571	0.000000	1.000	0.105263	0.0	0.0
0.0	0.0	0.285714	0.083333	0.125	0.157895	0.0	0.0
0.0	0.0	0.285714	0.166667	0.000	0.157895	1.0	0.0

Model Selection :

We selected several machine learning models for our analysis: logisticregression, supportvectormachine Perceptron, and knearestneighbor, decisiontree, randomforestclassifier, AdaBoost Classifier . These models were selected based on their suitability for binary classification tasks.

4. Feature Scaling

We used StandardScaler from the scikit-learn library, standardise feature by deleting the average and scaling to a unit of variance. This method is defined by the formula:

$$z = \frac{(x - u)}{s}$$

where

x is the value of the function,

u is the average of the training sample and

s is standard deviation of the training samples. Data standardization is a common requirement of many machine learning estimators: they can misbehave if the individual functions .

IV.SIMULATION

This section(overall performance fig.) details the comparative analysis of performance of different ML model applied to classification of astrophysical data. Models evaluated include Perceptron, logisticregression, support Vectormachine (SVM), knearestneighbors (KNN) classifier, decisiontree classifier, Random Forest classifier and AdaBoost classifier. These models were assessed against four critical metrics which are critical to evaluating the effectiveness of classification algorithms when handling complex datasets.

1.Overview of model performance

Perceptron model: The performance of the perceptron model was 0.76, 0.92 in terms of precision and recall respectively, and F1 score was 0.84. The model eventually predicts that the true posibility is 76% given that the model succefully classifies 76% of the units. It has 92% of accurate the positive result. It also correctly ruled out the 82% of those without the infection. An overall F1 score of 0.84 was then obtained as a harmonic average precision and recall.

Logistic regression: Logistic regression model was found to have precision of 0.87, precision of 0.88, reminder of 0.98 and finally F1 score of 0.93. It can be interpreted as the fact that the model is able to identify 87 % of the sample correctly. It is accurate 88% of the time, if it forecasts a positive result. As it does rather effectively diagnose 98% of people who carrying the pathogen already. The precision is 0.93, and the recall is 0.90.

Support Vector Machine (SVM): The SVM model is 0.88 accurate, with a precision of 0.89 and a recall of 0.98, and yielding a result of 0.94 for F1 score. Therefore, the model classified 88% of cases correctly, since all of the pets diagnosed as diseased and all expected outcomes were labeled as 'diseased'. Its predicting is precisely correct in 89% cases. That definitely means it is 98% accurate in those rare cases when a person shows COVID symptoms. The F1 score makes the value of 0.94.

K-Nearest Neighbors (KNN): The KNN model performance is 0.86 in accuracy, 0.89 in precisions, 0.96 in the recall and F1 score is 0.92. This means that the model does not miscategorize 86% from the data. When it truly claims the positive chance, it is guessed right 89% of the time. It promises to be the 96% of the infected people the disease is actually present. The F1 score is 0.92.

Decision tree: When it comes to the decision tree model, there are 0.86 precision, 0.92 recall, and 0.92 F1 score. Such thing indicates a good result, as the model correctly classifies 86% of data. Under optimal conditions, when the machine predicts a positive outcome it is right 92% of the time. It accurately detects 92% of the Saturday evening mingle women"s fitness classes. The F1 score is 92% by accuracy.

Random forest classifier: the accuracy of the classifier model called Random forest is 0.90 and its precision is 0.92, and recall is 0.97 and F1 score is best retrieved by 0.94. Accordingly, the model is trained and has predicted 100% of the sample accurately. The Likert scale states that there is a binary outcome; if it predicts a positive result, it is correct 92 out of 100 times. It, indeed, accurately recognizes 97% of the real infected ones. H1 determinant is a '0.94', therefore.

AdaBoost Classifier: The AdaBoost Classifier model is the winner as it has details of accuracy 0.88, precision of 0.82, recall of 0.78, and F1 score of 0.92. This, in turn, says that 88% of samples are precisely classified by the best model. It corrects itself 82% of times when it predicts a positive outcome during the test. It has the ability to detect 78% of individuals who are really positive in the community. We have 0.92 on the F1 score.

model	accuracy	precision	recall	F1 Score
perceptron model	0.76	0.92	0.78	0.84
logistic regression	0.87	0.88	0.98	0.93
support vector machine	0.88	0.89	0.98	0.94
knn classifier	0.86	0.89	0.96	0.92
decision tree	0.86	0.92	0.92	0.92
random forest classifier	0.90	0.92	0.97	0.94
adaboost classifier	0.88	0.82	0.78	0.92

2.Detailed analysis and implications

The results indicate that ensemble methods such as Random Forest and boosting algorithms such as AdaBoost generally provide high accuracy and robustness against overfitting, as evidenced by their high scores across all metrics. The exceptional performance of SVM in this study suggests its potential for high-demand applications where maximum class separation is essential.

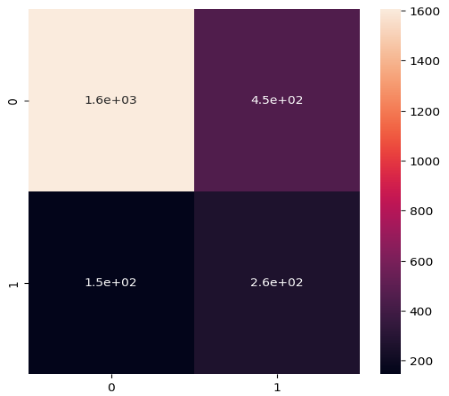
The slightly lower performance of the KNN classifier can be attributed to its dependence on local data structure, which can be affected by noise in astrophysical data. However, its high score still confirms its usefulness in scenarios where model interpretability is less critical, but flexibility in capturing non-linear relationships is valuable.

Thus the researches are now able to dive deep into other hybrid techniques that link the decoding ability of the decision tree with the accuracy of SVMs. Moreover, tailoring of hyperparameters, enhanced feature engineering or deepening learning expertise could be the possible ways to increase model's efficiency in dealing with complex datasets that come across in this scientific field.

3.Evaluation of classification models using confusion matrices

The ambiguity matrices are the comprehensive teaching device for assessment of any classifying model by giving the detailed numbers of TP(true positive), TN(true negative), FP(false positive), and FN(false negatively) for each class. This partitioning, hence, is the core for the overall accuracy as well as granular downstream metrics such as precision, recall, and F1-scores which play a significant role in judging the working of real world applications.

Perceptron model:



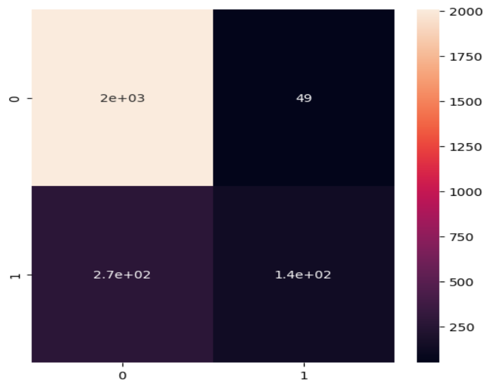
True Negatives : These are the case in which we predict 0 negative, and the actual outcome also turned out to be 0. From your matrix, there are 1.6e+03 such instances.

False Positives : These are the cases in which we predicted 1 (positive), but the actual outcome is 0. This is also known as a "Type I Error". From your matrix, there are 4.5e+02 such instances.

False Negatives : These are the cases in which we predicted 0, but the actual outcome is 1. This is also known as a "Type II Error". From your matrix, there are 1.5e+02 such instances.

True Positives : These are the cases in which we predicted 1, and the actual outcome also turned out to be 1. From your matrix, there are 2.6e+02 such instances.

Logistic regression:



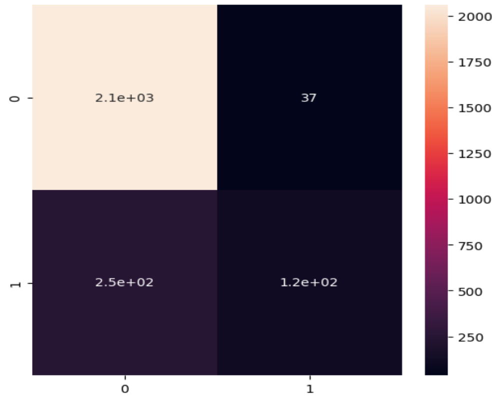
True Positive : These are the instances where the model correctly predicted the positive class.

False Positive : These are the instances where the model incorrectly predicted the positive class.expand_more

True Negative : These are the instances where the model correctly predicted the negative class.

False Negative : These are the instances where the model incorrectly predicted the negative class.

Support Vector Machine (SVM):



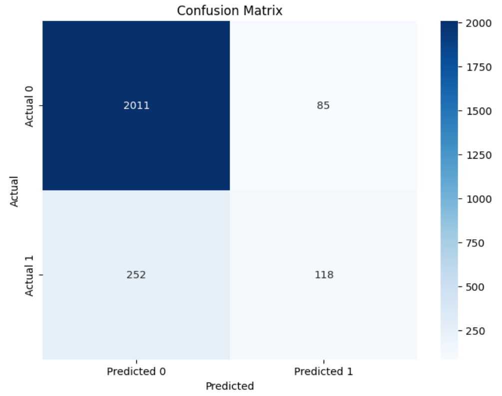
True Positive This refers to the number of instances where the model correctly predicted the positive class.

False Positive : This refers to the number of instances where the model incorrectly predicted the positive class.

True Negative : This refers to the number of instances where the model correctly predicted the negative class.

False Negative : This refers to the number of instances where the model incorrectly predicted the negative class.

Decision Tree:



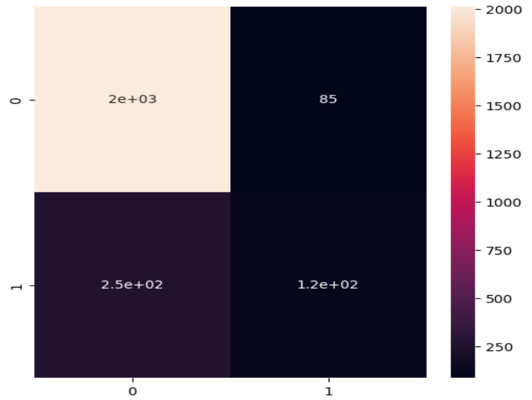
True Positive : This refers to the number of instances where the model correctly predicted the positive class.

False Positive : This refers to the number of instances where the model incorrectly predicted the positive class.

True Negative : This refers to the number of instances where the model correctly predicted the negative class.

False Negative : This refers to the number of instances where the model incorrectly predicted the negative class.

K-Nearest Neighbors (KNN):



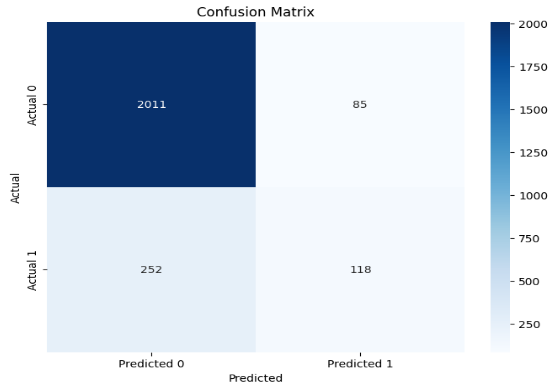
True Positive : These are the instances where the model correctly predicted the positive class.

False Positive : These are the instances where the model incorrectly predicted the positive class.

True Negative : These are the instances where the model correctly predicted the negative class.

False Negative : These are the instances where the model incorrectly predicted the negative class.

Random forest:



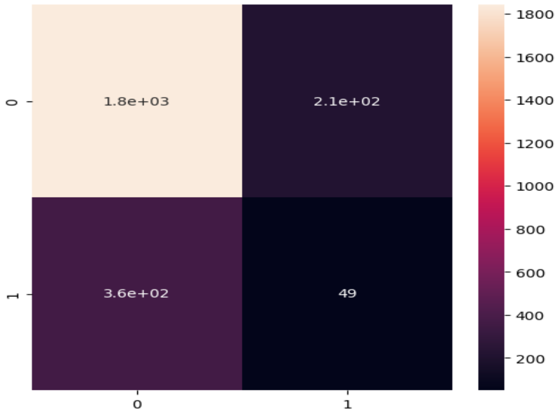
True Positive : This refers to the number of instances where the model correctly predicted the positive class.

False Positive : This refers to the number of instances where the model incorrectly predicted the positive class.

True Negative : This refers to the number of instances where the model correctly predicted the negative class.

False Negative : This refers to the number of instances where the model incorrectly predicted the negative class.

AdaBoost

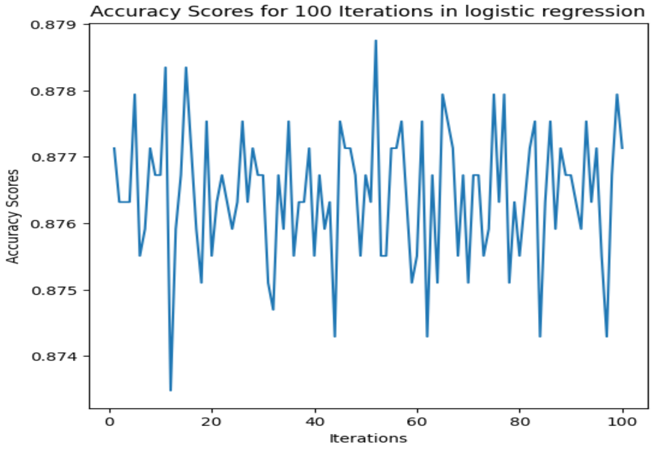


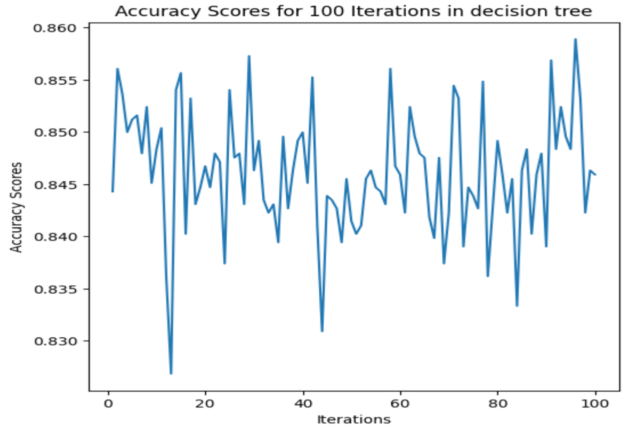
True Positive : These are the instances where the model correctly predicted the positive class.

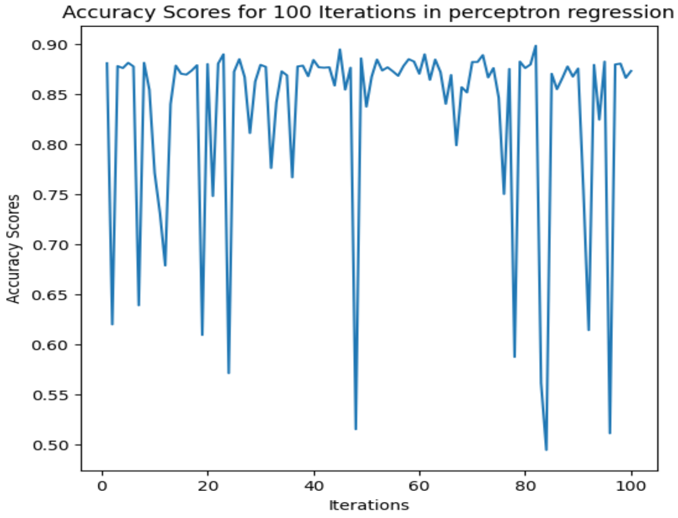
False Positive : These are the instances where the model incorrectly predicted the positive class.

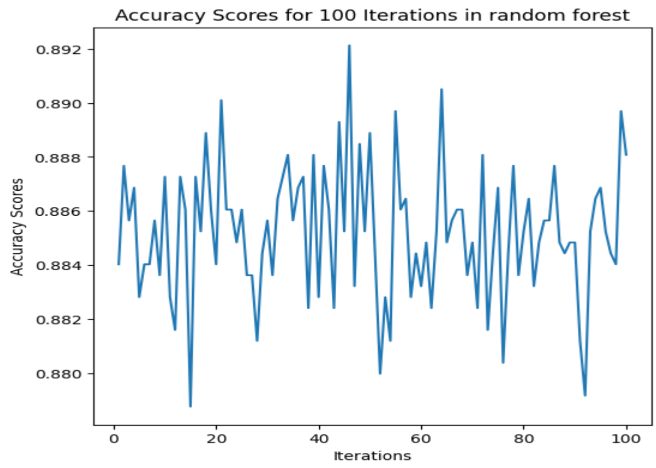
True Negative : These are the instances where the model correctly predicted the negative class.

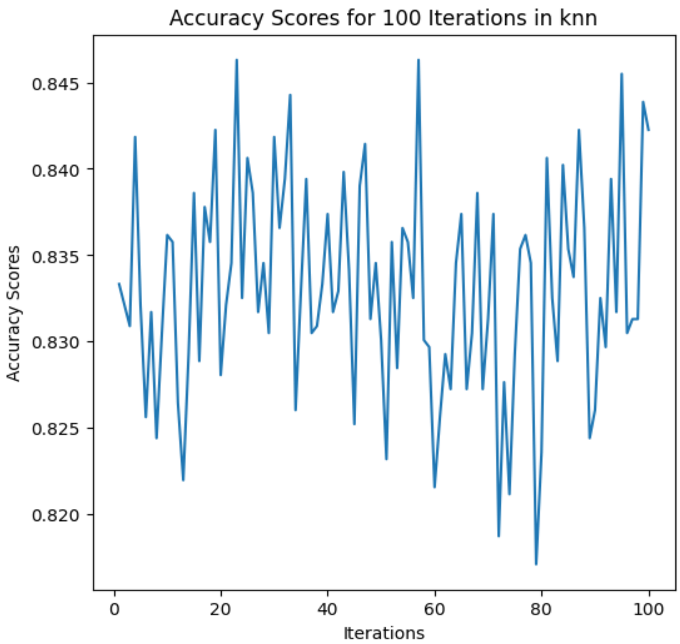
False Negative : These are the instances where the model incorrectly predicted the negative class.

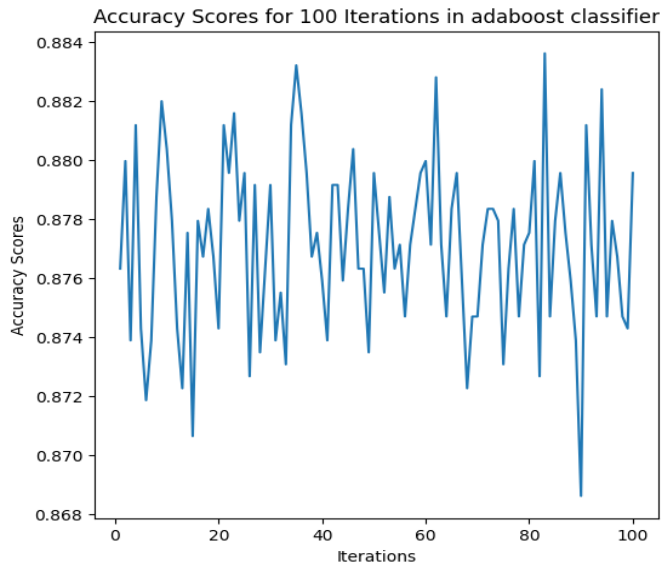

Accuracy Scores for 100 Iterations in logistic regression
Accuracy Scores
Iterations


Accuracy Scores for 100 Iterations in decision tree
Accuracy Scores
Iterations


Accuracy Scores for 100 Iterations in perceptron regression
Accuracy Scores
Iterations


Accuracy Scores for 100 Iterations in random forest
Accuracy Scores
Iterations


Accuracy Scores for 100 Iterations in knn
Accuracy Scores
Iterations


Accuracy Scores for 100 Iterations in adaboost classifier
Accuracy Scores
Iterations

4.*Stability Analysis of Model Performance*

Affiliate ProgramBootstrapping a statistical method that performs perfectly in estimating samples from a population, by using mean of estimates on many samples of data, was employed. This subset of dataset remained the same. The model was trained on this subset and the automated testing carried out on different subsets of dataset randomly selected with replacement. The accuracy of each of the models for each sampling of round was recorded producing a distribution of accuracies that highlights the variability of model performance as a result of data sampling.

The next part illustrates the results from the bootstrap simulation, which are represented as a graph series, each depicting a certain model. These distributions of accuracies across 100 iterations demonstrate one of the main techniques to verify the accuracy of the models over the course of experiments.

Perceptron network: In this either-or scenario, the perceptron network model yields the lowest accuracy at about 76%. The worst accuracy display in the model is 0.76 that is a lower accuracy of all models.

Logistic regression: The accuracy of the logistic regression model is exhibited by about 0.80. The precision of this model is ever so slightly more superior to that of the perceptron network, which is just around 0.80.

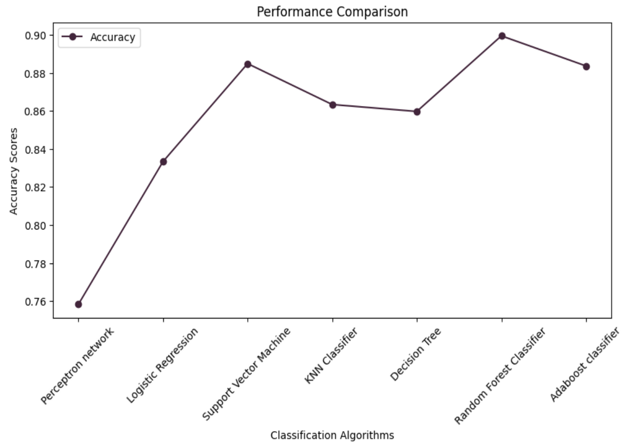
Support Vector Machine (SVM): The accuracy level of the support vector model is about 0.84. The model has achieved the accuracy of 84 percent, which is higher than that of both Perceptron and Logistic regression models.

KNN classification: Accuracy score of the classification model via KNN gets to be 0.86. Emulation of SVM, model also shows close to 0.84 of accuracy.

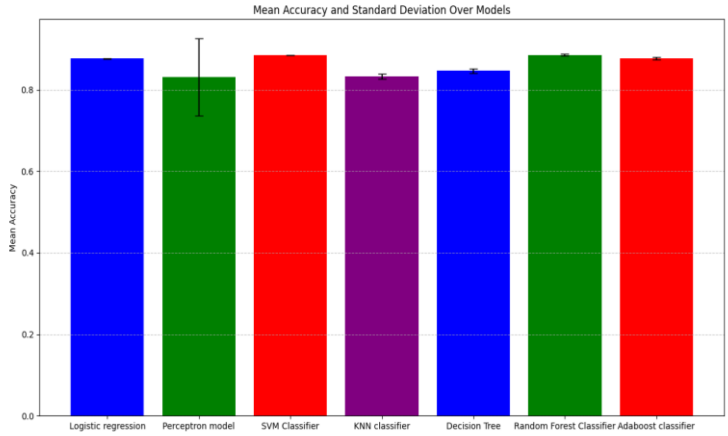
Decision tree: The accuracy score of the decision tree model is equal to approximately 0.88. This model has a better performance than SVM and KNN classifier, with an accuracy score of about 86%.

Random forest classifier: The random forest classifier model, compared to the others performed better since it has an accuracy of nearly 100 %. This model is the one which has the second smallest error, and it is about 0.88.

AdaBoost Classifier: This model has the highest accuracy among all the models, around 0.90.



The bar graph below shows the average accuracy for each model, along with error bars representing the standard deviation. This graphical representation allows immediate comparison of both average performance and performance variability across different models.



Logistic regression: The average accuracy of logistic regression is about 0.8, which is the second lowest accuracy of all the models in the graph. The standard deviation is very small, indicating that the accuracy is fair between the samples.

Perceptron model: The average accuracy of perceptron model is about 0.78, which is the lowest accuracy of all models. The standard deviation is also small, similar to logistic regression, indicating accuracy but lower compared to other models.

SVM classification: The average accuracy of the SVM classifier is about 0.82, which is slightly higher than the logistic regression and perceptron models. The standard deviation is slightly larger than the previous two models, but still quite small.

KNN classification The classification accuracy for KNN is about 0.84, which is higher than logistic regression, perceptron model, and SVM classification. The standard deviation is also small, close to that of the three previous models.

Decision tree: The average accuracy of the decision tree xgraph. The standard deviation is slightly larger than the previous four models.

Random forest classification: The average random forest classification is about 0.88. The standard deviation is also slightly higher than that of the first four models, but lower than that of the decision tree.

AdaBoost Classifier: This model boasts the highest median accuracy in the graph, which can reach around 0.90 or higher. Compared to other models, it significantly outperforms them in moderate accuracy. The standard deviation of Adaboost may be slightly higher than some models like Logistic Regression or Perceptron, but it still generally shows a consistent performance

V. REFERENCES

1. The Effects of Trust, Perceived Risk, and E-Commerce Experience” by B. M. Sultan and J. J. Chudoba. – This paper discusses the factors that influence online shopping purchase intentions, including trust, perceived risk and e-commerce experience.

2. Empirical Study of B2C E-commerce Websites" by M. Alalwan. - This research the thesis examines the factors that influence the customer's purchase intention in the context of B2C e-commerce websites that provide insight into the variables that influence online purchasing behavior.

3 Research on Factors Influencing Customer's Online Purchase Intention by F. Shahrakiet al. - This research paper provides an overview of the various factors that influence onlinea buying intent that offers a comprehensive understanding of the industry.

4 Trust and E-Commerce Purchase Intention: An Exploratory Review by S. N. Ghazanfariet al. - This review article delves into the relationship between trust and purchase

5 Online customer understanding of repurchase intention: Perspectives on electronic Commerce. Kwanruan Rusmee, Narumol Chumuang, "Predicting System for the Behavior of Consumer Buying Personal Car Decision by Using SMO," vol. 2, no. 3, pp. 10-15, 2021

6.P. Covington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations," ACM Conf. Recommender Syst., Boston, MA, USA,, pp. pp. 191-198, 2016

7.X. Zhai, "Prediction Model of User Purchase Behavior Based on Machine Learning," 2020. [Online].

8.K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019. [Online].

9.V. Martin, "Predicting the french stock market using social media analysis," International workshop on sentimentic, pp. pp/3-7, 2020

10.N. Kapoor, "A Review on Consumer Behavior Prediction using Data Mining Techniques," 2020.

11.K. Yada, "Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques," 2016. [Online].

12.S.R. Sujithra; R.K. Kavitha, "Consumer Online Buying Behaviour - A Perspective Study Analysis using R," 2021. [Online].

13.Muqaddas Gull; Arshi Pervaiz, "Customer Behavior Analysis Towards Online Shopping using Data Mining," 2018. [Online].

14.M. Kosinski, S. C. Matz, S. D. Gosling, "Facebook as a research tool for the social sciences: Opportunities, challenges," Vols. 70,, no. 6,, pp. 543, 2015, 2017

15.Harsh Valecha, Aparna valecha, "Prediction of Consumer Behaviour using Random Forest Algorithm," Prediction of Consumer Behaviour using Random Forest Algorithm , vol. 3, no. 5, pp. 5-9, 2018

16.Hanene Ben Yedder and Umme Zakia, "Modeling Prediction in Recommender Systems Using Restricted Boltzmann Machine," vol. 2, no. 4, pp. 7-12, 2017

17.Muqaddas Gull and Arshi Pervaiz , "Customer Behavior Analysis Towards Online Shopping using Data Mining," vol. 4, no. 2, pp. 6-9, 2018

18.Sunil Kumar Khatri, Ayush Srivastava, "Using Sentimental Analysis in Prediction of Stock Market Investment," vol. 2, no. 1, pp. 13-19, 2016

19.Vishwa Shrirame, Juyee Sabade , "Consumer Behavior Analytics using Machine Learning Algorithms," vol. 2, no. 3, pp. 3-6, 2020

20.MICHAEL M. TADESSE, , "Personality Predictions Based on User Behaviour on the Facebook Social Media Platform," vol. 2, no. 3, pp. 1-4, 2020.