MACHINE LEARNING PROJECT

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted by:

Jigisha Chopra 102003650

Yatin Goyal 102003655

Submitted to:

Dr. Harpreet Singh

Online shopper's intention dataset

Introduction

We are addressing the online shopper's intention dataset here. It is used to predict a customer's behaviour in online shopping websites, for Key performance indicators and marketing analysis. Here, we analyse the behaviour of customers as they browse through the pages to predict if they will make a purchase or not. How do we know if a customer is going to shop or walk away? Understanding the customers is crucial to any seller/store/online platform. This understanding can be important in convincing a customer who is just browsing to buy a product. In offline stores, the inferences derived influence the placement of objects in the store. When the same experience is translated to an online store, the sequence of web pages browsed to reach a product becomes important.

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn. By using statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, impacting growth. They will be required to help identify the most relevant business questions and the data to answer them. Like for this project it helps us predict that how do we know if a customer is going to shop or walk away? Necessary changes to be made to increase revenue? Therefore, Machine learning algorithms are typically created using frameworks that accelerate solution development.

In further sections we will discuss about the background of the problem, dataset description, architecture of the framework, brief information about the algorithms used and the conclusion derived. Data science life cycle is used to solve the problem. Steps like data collection, data preparation, EDA, modelling and evaluation etc are used.

Background

1. **Online repatronage intention: an empirical study among Malaysian experienced online shoppers**
   The purpose of this paper is to sketch and determine the impact of perceived usefulness (PU), perceived ease of use (PEOU), perceived value (PV), trust (TRT), perceived risk (PR), privacy concern (PC), internet literacy (IL), satisfaction (SAT) on online repatronage intention (ORI) among Malaysian experienced online shoppers.
   The statistical analyses support the relationships between PU, PV, TRT and SAT with ORI while the relationships between PEOU, PR, PC and IL with ORI were rejected in which all the factors affecting ORI occur similarly across the study sample.

   Reference:

Rezaei, S., Amin, M. and Khairuzzaman Wan Ismail, W. (2014), "Online repatronage intention: an empirical study among Malaysian experienced online shoppers", *International Journal of Retail & Distribution Management*, Vol. 42 No. 5, pp. 390-421. https://doi.org/10.1108/IJRDM-03-2012-0026

2. **Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural network**

we propose a real-time online shopper behavior analysis system consisting of two modules which simultaneously predicts the visitor's shopping intent and Web site abandonment likelihood. In the first module, we predict the purchasing intention of the visitor using aggregated pageview data kept track during the visit along with some session and user information. The extracted features are fed to random forest (RF), support vector machines (SVMs), and multilayer perceptron (MLP) classifiers as input. We use oversampling and feature selection preprocessing steps to improve the performance and scalability of the classifiers.

Reference:

Carmona CJ, Ramírez-Gallego S, Torres F, Bernal E, del Jesús MJ, García S (2012) Web usage mining to improve the design of an e-commerce website: OrOliveSur. com. Expert Syst Appl 39(12):11243–11249

3. **The effect of product review balance and volume on online Shoppers' risk perception and purchase intention**

Study validates some findings in econometric studies pertaining to review balance and volume, with some differences regarding the effects of review volume. study enriches researchers' understanding of risk perception and uncertainty in e-commerce. study reveals that the relationship between perceived uncertainty and purchase intention becomes insignificant in the presence of attitude toward purchasing.

Reference:

J.A. Chevalier, D. Mayzlin

The effect of word of mouth on sales: online book reviews

Journal of Marketing Research, 43 (2006), pp. 345-354

4. **Effects of inertia and satisfaction in female online shoppers on repeat-purchase intention: The moderating roles of word-of-mouth and alternative attraction**

With the prevalence of the internet, whether various interactive relationship building between online channel and consumers may lead or not to profit has been paid much attention by researchers and practitioners. The study results indicate that both consumer inertia and satisfaction positively influence repeat-purchase intention, and that consumer inertia is more influential than satisfaction; moreover, positive word-of-mouth negatively moderates the relationship between consumer inertia and repeat-purchase intention, but positively moderates that between satisfaction and repeat-purchase intention; finally, alternative attraction does not moderate any of the above relationships significantly.

Reference:

Kuo, Y., Hu, T. and Yang, S. (2013), "Effects of inertia and satisfaction in female online shoppers on repeat-purchase intention: The moderating roles of word-of-mouth and alternative attraction", *Managing Service Quality: An International Journal*, Vol. 23 No. 3, pp. 168-187. https://doi.org/10.1108/09604521311312219

5. Real-Time Prediction of Online Shoppers’ Purchasing Intention Using Random Forest

we suggest a real-time online shopper behavior prediction system which predicts the visitor’s shopping intent as soon as the website is visited. To do that, we rely on session and visitor information and we investigate naïve Bayes classifier, C4.5 decision tree and random forest. Furthermore, we use oversampling to improve the performance and the scalability of each classifier. The results show that random forest produces significantly higher accuracy and F1 Score than the compared techniques.

Reference:

Sakar, C.O., Polat, S.O., Katircioglu, M., Kastro, Y.: Real-time prediction of online shoppers’ purchasing intention using multilayer perceptron and LSTM recurrent neural networks. Neural Comput. Appl. **31**(10), 6893–6908 (2019). https://doi.org/10.1007/s00521-018-3523-0

6. Do Vendor Cues Influence Purchase Intention of Online Shoppers? An Empirical Study Using S-O-R Framework

The purpose of the current research is to understand the influence of vendor cues like brand reputation. Brand familiarity and offline presence on trust and attitude of online shoppers and consequently on online purchase intention. Data was collected through a web based survey. The findings of the study reported that vendor offline cues have a strong and positive impact on the online purchase intentions of the shoppers. Further, this study also contributed by proving that the trust has a strong relationship with purchase intention as compare to attitude.

Reference:

Adaval, R. 2003. How good gets better and bad gets worse: Understanding the impact of affect on evaluation of known brands. *Journal of Consumer Research* 30 (3):352–67. doi:10.1086/378614 [Crossref], [Web of Science ®]

7. Fresh Produce E-Commerce and Online Shoppers’ Purchase Intention

The development of the Internet has provided many opportunities for electronic commerce, and many e-commerce companies like Alibaba have achieved great success. Fresh produce industry has also attempted to step into e-commerce during the past decade. It is important for e-commerce to understand customers’ demands in this new market in order to make profits. In this research we conducted a market survey to investigate the market situation of Chinese fresh produce e-commerce. Consumer attitudes and behaviors toward online shopping for fresh fruits were evaluated. A logit model was used to identify potential factors that may have impact on consumers’ purchase intention. Results show that women are more likely than men to shop online; other factors such as influence from friends, income, product quality, food labels, packaging, and payment security can also affect online shoppers’ purchase intention.

Reference:

An, L., Han, Y., & Tong, L. (2016, May). Study on the factors of online shopping intention for fresh agricultural products based on UTAUT2. In *The 2nd Information technology and mechatronics engineering conference (ITOEC 2016)*. Atlantis Press. https://doi.org/10.2991/itoec-16.2016.57 [Crossref]

About the dataset

Data Collection

We would be using Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning repository for our task/project/problem)

Feature	Explanation	Measurement	Range
Administrative	Pages visited by visitor about account management	Numeric	[0, …, 27]
Admin_Duration	Seconds spent by visitor on account management related pages	Numeric	[0, …, 3398]
Informational	Informational pages visited by visitor	Numeric	[0, …, 24]
Info_Duration	Seconds spent by visitor on informational pages	Numeric	[0, …, 2549]
ProductRelated	Pages visited by visitor about products	Numeric	[0, …, 705]
ProductRelated_Dur	Seconds spent by visitor on product related pages	Numeric	[0, …, 63973]
BounceRates	Average bounce rate value of the pages visited by visitor	Numeric	[0, …, 0.2]
ExitRates	Average exit rate value of pages visited by visitor	Numeric	[0, …, 0.2]
PageValues	Average page value of the pages	Numeric	[0, …, 361]
SpecialDay	Closeness of the site visiting time to a special day	Numeric	[0, …, 1]
Month	Month value of the visit date	Months	12
OperatingSystems	Operating system of the visitor	Numeric	8
Browser	Browser of the visitor	Numeric	13
Region	Geographic region from which the session has been started by the visitor	Numeric	9
TrafficType	Traffic source (e.g., banner, SMS, direct)	Numeric	20
VisitorType	Visitor type (e.g., new, returning, other)	String	3
Weekend	Boolean value indicating whether the date of the visit is weekend	Boolean	2

Revenue	Class label: whether the visit has been finalised with a transaction	Boolean	2

Data Preparation and Cleaning

- Favourably are no missing values in the data set being used. There are techniques to handle such situations with the help of Pandas and NumPy libraries of python.
- By observing the data, we get to know that there are 17 independent features and one dependent variable i.e Revenue. To perform the visualization of data, it is necessary to convert all features to numerical form. We can transform it by using the Sklearn library.
- The month column needs to be one-hot encoded with all the 12 months in the count.
- The Weekend and Revenue columns having data type boolean are converted to integer.
- The columns Operating Systems, Browser, Region, TrafficType, and VisitorType have values that do not have any range dependency. They are one hot encoded using the get_dummies() function of the pandas library. The drop_first parameter is set to True. This means that if there are n categories in the column, n-1 columns are returned instead of n.

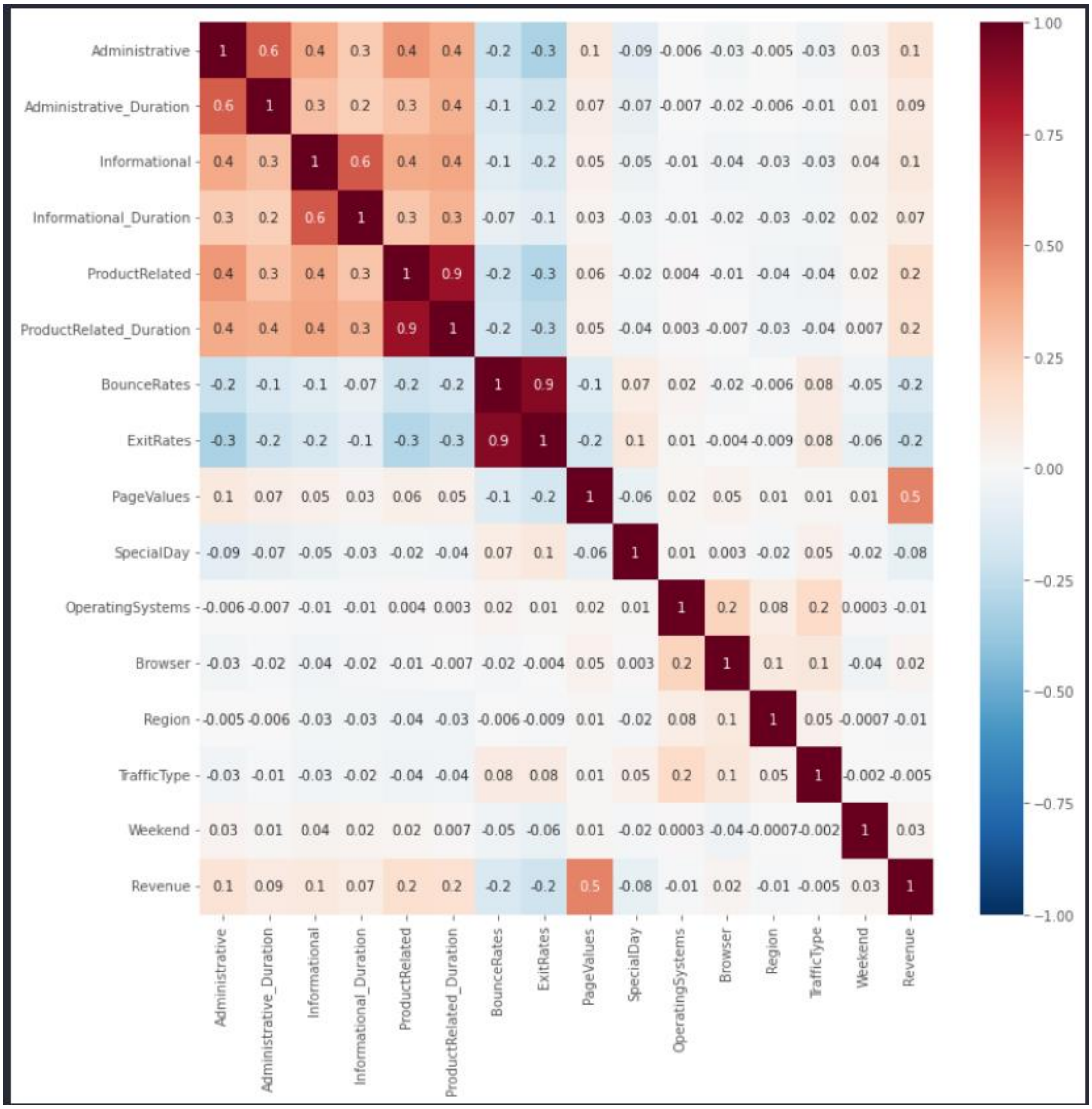
Train-validation-test split

The train_test_split function of the sklearn.model_selection module is used. Ratio of 70-15-15 used to split data frame into train – validation- test sets.

Standardisation

The standard deviation of attribute values in the dataset is not the same across all of them. This may result in certain attributes being weighted higher than others. The values across all attributes are scaled to have mean = 0 and standard deviation = 1 with respect to the particular columns. The Standard Scaler function of the sklearn. pre-processing module is used to implement this concept. The instance is first fit on the training data and used to transform the train, validation, and test data.

Feature selection



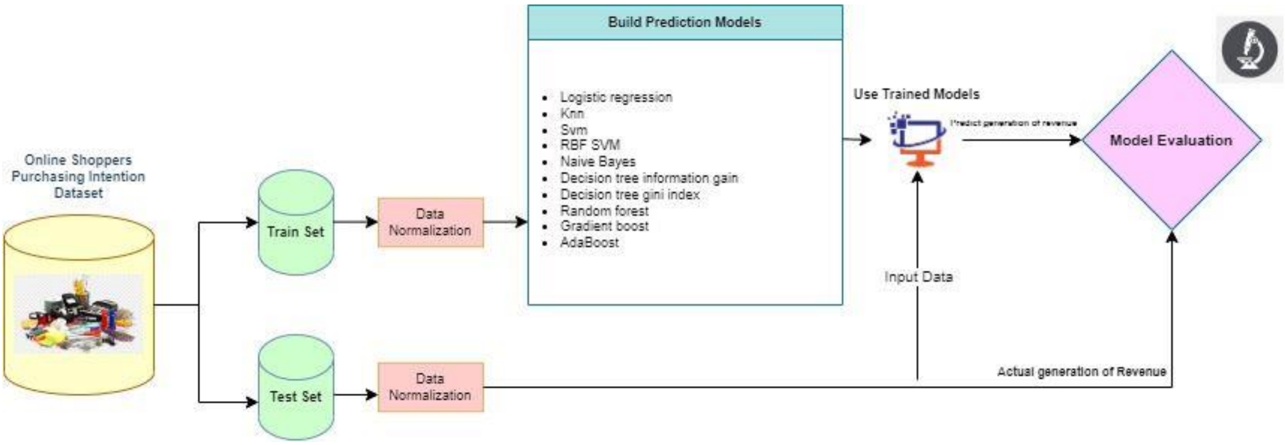
	Administrative	Administrative_Duration	Informational	Informational_Duration	ProductRelated	ProductRelated_Duration	BounceRates	ExitRates	PageValues	SpecialDay	OperatingSystems	Browser	Region	TrafficType	Weekend	Revenue
Administrative	1	0.6	0.4	0.3	0.4	0.4	-0.2	-0.3	0.1	-0.09	-0.006	-0.03	-0.005	-0.03	0.03	0.1
Administrative_Duration	0.6	1	0.3	0.2	0.3	0.4	-0.1	-0.2	0.07	-0.07	-0.007	-0.02	-0.006	-0.01	0.01	0.09
Informational	0.4	0.3	1	0.6	0.4	0.4	-0.1	-0.2	0.05	-0.05	-0.01	-0.04	-0.03	-0.03	0.04	0.1
Informational_Duration	0.3	0.2	0.6	1	0.3	0.3	-0.07	-0.1	0.03	-0.03	-0.01	-0.02	-0.03	-0.02	0.02	0.07
ProductRelated	0.4	0.3	0.4	0.3	1	0.9	-0.2	-0.3	0.06	-0.02	0.004	-0.01	-0.04	-0.04	0.02	0.2
ProductRelated_Duration	0.4	0.4	0.4	0.3	0.9	1	-0.2	-0.3	0.05	-0.04	0.003	-0.007	-0.03	-0.04	0.007	0.2
BounceRates	-0.2	-0.1	-0.1	-0.07	-0.2	-0.2	1	0.9	-0.1	0.07	0.02	-0.02	-0.006	0.08	-0.05	-0.2
ExitRates	-0.3	-0.2	-0.2	-0.1	-0.3	-0.3	0.9	1	-0.2	0.1	0.01	-0.004	-0.009	0.08	-0.06	-0.2
PageValues	0.1	0.07	0.05	0.03	0.06	0.05	-0.1	-0.2	1	-0.06	0.02	0.05	0.01	0.01	0.01	0.5
SpecialDay	-0.09	-0.07	-0.05	-0.03	-0.02	-0.04	0.07	0.1	-0.06	1	0.01	0.003	-0.02	0.05	-0.02	-0.08
OperatingSystems	-0.006	-0.007	-0.01	-0.01	0.004	0.003	0.02	0.01	0.02	0.01	1	0.2	0.08	0.2	0.0003	-0.01
Browser	-0.03	-0.02	-0.04	-0.02	-0.01	-0.007	-0.02	-0.004	0.05	0.003	0.2	1	0.1	0.1	-0.04	0.02
Region	-0.005	-0.006	-0.03	-0.03	-0.04	-0.03	-0.006	-0.009	0.01	-0.02	0.08	0.1	1	0.05	-0.0007	-0.01
TrafficType	-0.03	-0.01	-0.03	-0.02	-0.04	-0.04	0.08	0.08	0.01	0.05	0.2	0.1	0.05	1	-0.002	-0.005
Weekend	0.03	0.01	0.04	0.02	0.02	0.007	-0.05	-0.06	0.01	-0.02	0.0003	-0.04	-0.0007	-0.002	1	0.03
Revenue	0.1	0.09	0.1	0.07	0.2	0.2	-0.2	-0.2	0.5	-0.08	-0.01	0.02	-0.01	-0.005	0.03	1

From the above heatmap, we observe the following:

- In general, there is very little correlation among the different features in our dataset.
- The very few cases of high correlation (|corr| >= 0.7) are:
  - BounceRates & ExitRates (0.9).
  - ProductRelated & ProductRelated_Duration (0.9).
- Moderate Correlations (0.3 < |corr| < 0.7):
  - Among the following features: Administrative, Administrative_Duration, Informational, Informational_Duration, ProductRelated, and ProductRelated_Duration.
  - Also, between PageValues and Revenue.

Hence, there is no need of feature extraction.

Architecture

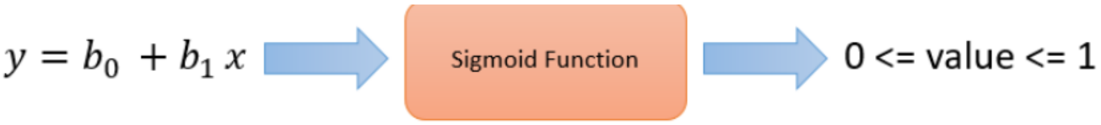


First, we understand the online shoppers intention prediction problem that what predictions need to be made and what market analysis needs to be performed. We would be using Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning repository for our task/project/problem).we remove any missing values in our data and convert all features to numerical form to perform visualization of data.we then perform data normalization which is a scaling technique in machine learning applied during data preparation to change or transform numeric values in dataset to lie in a certain range to use a common scale.we then split our data into training,test and validation set. We build and use 10 different learning classifiers (Naive Bayes, KNN, SVM, Logistic Regression, Random Forest, Gradient Boosting, Decision Tree and Adaboosting). These are tested and optimized until we have achieved the best classification performance. Model Evaluation is done based on parameters like specificity,recall,accuracy,error rate, TPR,FPR,AUC.confusion matrix is drawn for each algorithm and a common roc curve plotted to compare performance of algorithms.

Algorithms used

1. Logistic regression without regularisation

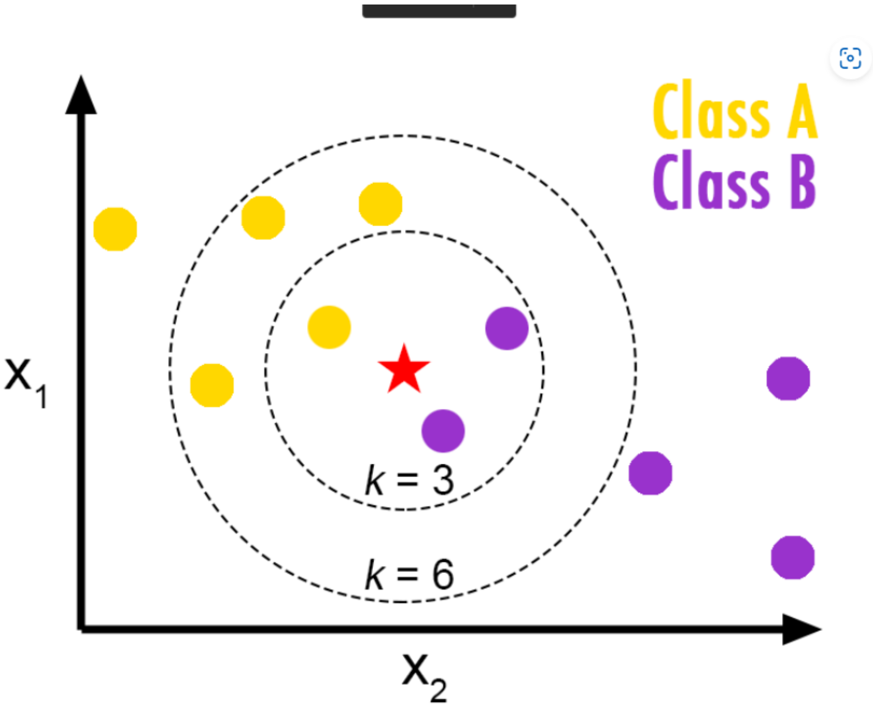
Logistic regression is a supervised classification algorithm which predicts the class or label based on predictor/ input variables (features).For example, by analyzing data, logistic regression can solve problems like finding if a customer will purchase a car or not, based on various factors like his/her age, salary, job profile, family type etc. In logistic regression, we find the appropriate coefficient values in a linear equation which give the least classification error. We provide this linear equation to sigmoid function. The sigmoid function maps the entire data into real numbers of range 0 to 1. By setting threshold for this output we can easily classify the input data.



$$\text{Inverse logit or sigmoid function} = logit^{-1} = \frac{1}{(1 + e^{-t})}$$

2. KNN Classifier

An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If *k* = 1, then the object is simply assigned to the class of that single nearest neighbor.



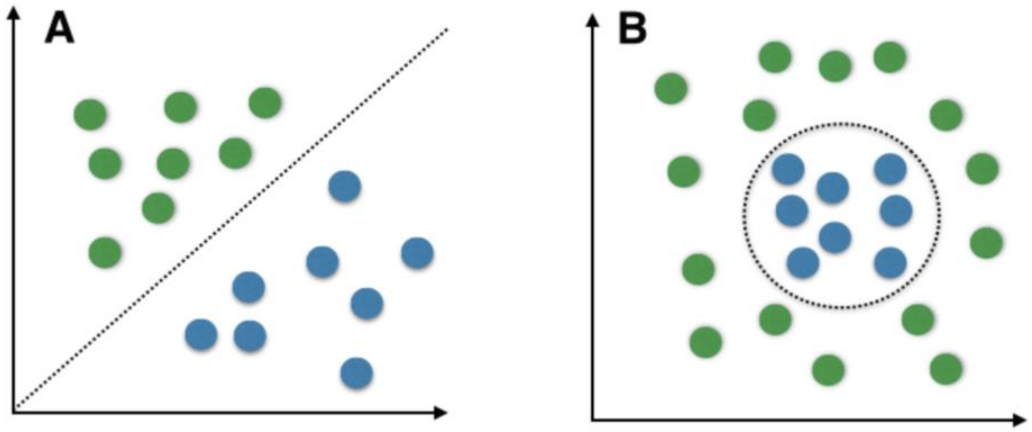
class of star would be B if k =3 and A if k = 6

3. Support Vector Machine

SVM finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line.we plot each data item in the dataset in an N-dimensional space, where N is the number of features/attributes in the data. Next, find the optimal hyperplane to separate the data. SVM can only perform binary classification (i.e., choose between two classes). However, there are various techniques to use for multi-class problems. To perform SVM on multi-class problems, we can create a binary classifier for each class of the data. The two results of each classifier will be :

- The data point belongs to that class OR
- The data point does not belong to that class.

For example, in a class of fruits, to perform multi-class classification, we can create a binary classifier for each fruit. For say, the 'mango' class, there will be a binary classifier to predict if it IS a mango OR it is NOT a mango. The classifier with the highest score is chosen as the output of the SVM. SVM works very well without any modifications for linearly separable data. **Linearly Separable Data** is any data that can be plotted in a graph and can be separated into classes using a straight line.



A: Linearly Separable Data B: Non-Linearly Separable Data

We use **Kernelized SVM** for non-linearly separable data. Say, we have some non-linearly separable data in one dimension. We can transform this data into two dimensions and the data will become linearly separable in two dimensions. This is done by mapping each 1-D data point to a corresponding 2-D ordered pair. So for any non-linearly separable data in any dimension, we can just map the data to a higher dimension and then make it linearly separable. This is a very powerful and general transformation. A **kernel** is nothing but a measure of similarity between data points. The **kernel function** in a kernelized SVM tells you, that given two data points in the original feature space, what the similarity is between the points in the newly transformed feature space. There are various kernel functions available we have used :

- **Radial Basis Function Kernel (RBF):** The similarity between two points in the transformed feature space is an exponentially decaying function of the distance between the vectors and the original input space as shown below. RBF is the default kernel used in SVM.

$$K(x, x') = exp(-\gamma||x - x'||)$$

4. Naive Bayes

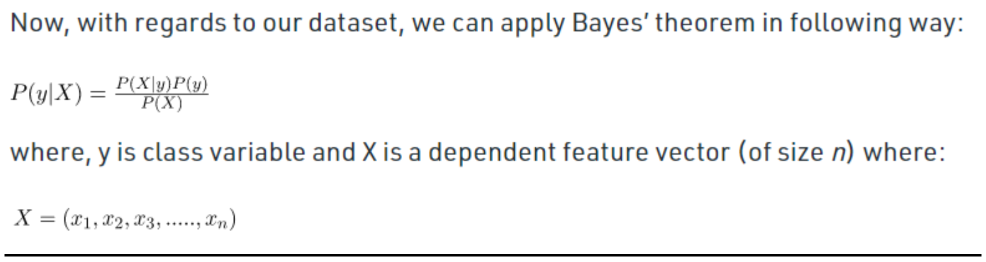
Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' Theorem.** The dataset is divided into two parts, namely, **feature matrix** and the **response vector**.

- Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of **dependent features**
- Response vector contains the value of **class variable** (prediction or output) for each row of the feature matrix

Assumption:

The fundamental Naive Bayes assumption is that each feature makes an:

- independent
- equal



Now, with regards to our dataset, we can apply Bayes' theorem in following way:

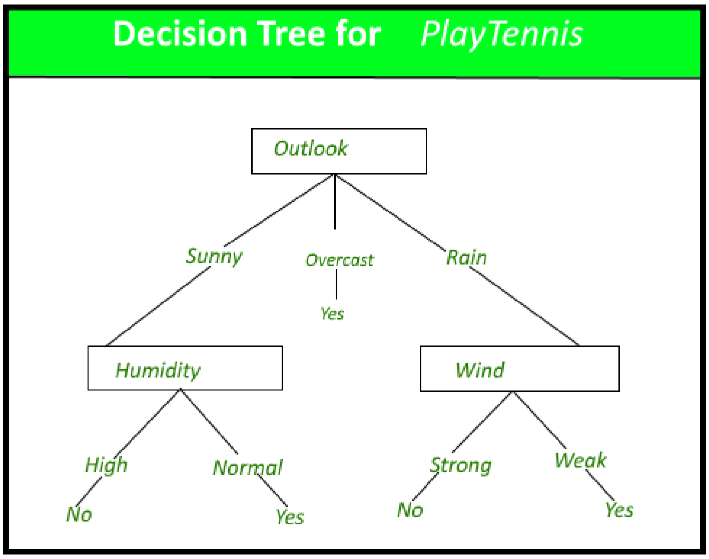
$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$

where, y is class variable and X is a dependent feature vector (of size *n*) where:

$X = (x_1, x_2, x_3, ..., x_n)$

5. Decision Tree

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.



In other words, we can say that the decision tree represents a disjunction of conjunctions of constraints on the attribute values of instances.

(Outlook = Sunny ^ Humidity = Normal) v (Outlook = Overcast) v (Outlook = Rain ^ Wind = Weak)

1. Information Gain
When we use a node in a decision tree to partition the training instances into smaller subsets the entropy changes. Information gain is a measure of this change in entropy.
Definition: Suppose S is a set of instances, A is an attribute, S_v is the subset of S with A = v, and Values (A) is the set of all possible values of A, then

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|}.Entropy(S_v)$$

Entropy
Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content.

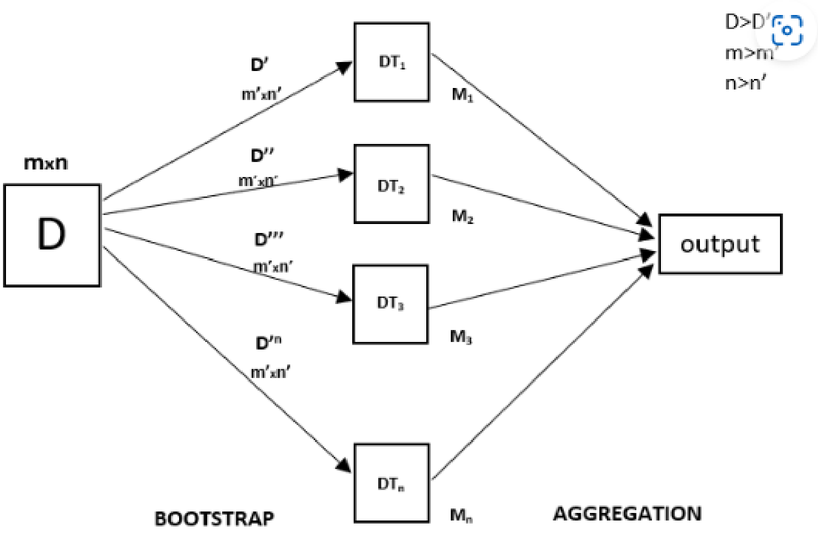
2. Gini Index is a metric to measure how often a randomly chosen element would be incorrectly identified.

It means an attribute with lower Gini index should be preferred.
Sklearn supports “Gini” criteria for Gini Index and by default, it takes “gini” value.

$$GiniIndex = 1 - \sum_j p_j^2$$

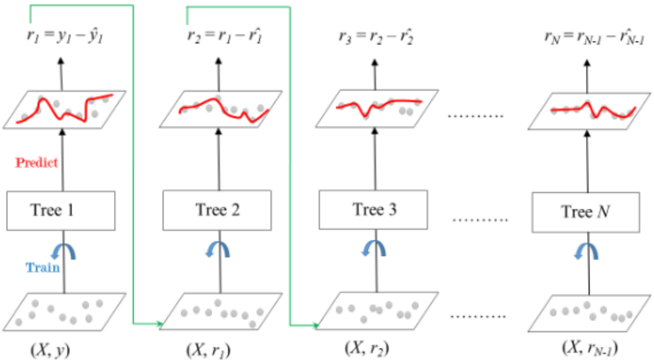
6. Random Forest

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as **bagging**. Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.



7. Gradient boost

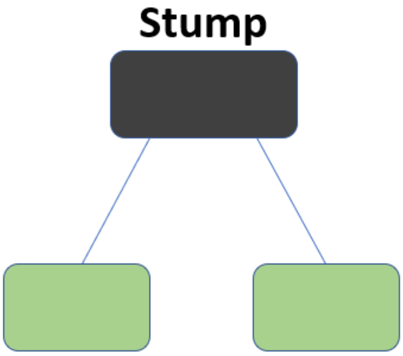
In gradient boosting, each predictor corrects its predecessor’s error. Each predictor is trained using the residual errors of predecessor as labels.
There is a technique called the **Gradient Boosted Trees** whose base learner is CART (Classification and Regression Trees).
The below d egression problems.



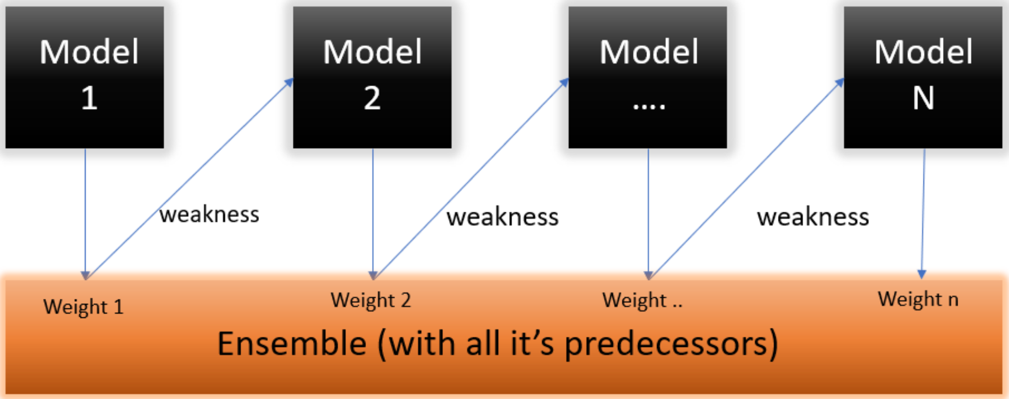
The ensemble consists of *N* trees. Tree1 is trained using the feature matrix *X* and the labels *y*. The predictions labelled *y1(hat)* are used to determine the training set residual errors *r1*. Tree2 is then trained using the feature matrix *X* and the residual errors *r1* of Tree1 as labels. The predicted results *r1(hat)* are then used to determine the residual *r2*. The process is repeated until all the *N* trees forming the ensemble are trained.

8. AdaBoost

AdaBoost also called Adaptive Boosting is a technique in Machine Learning used as an Ensemble Method. The most common algorithm used with AdaBoost is decision trees with one level that is Decision trees with only 1 split also called **Decision Stumps.**

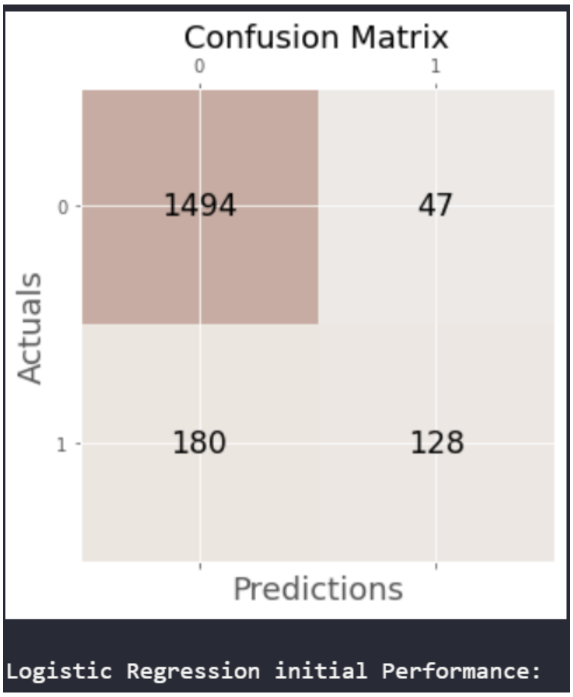


What this algorithm does is that it builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. Now all the points which have higher weights are given more importance in the next model. It will keep training models until and unless a lower error is received.

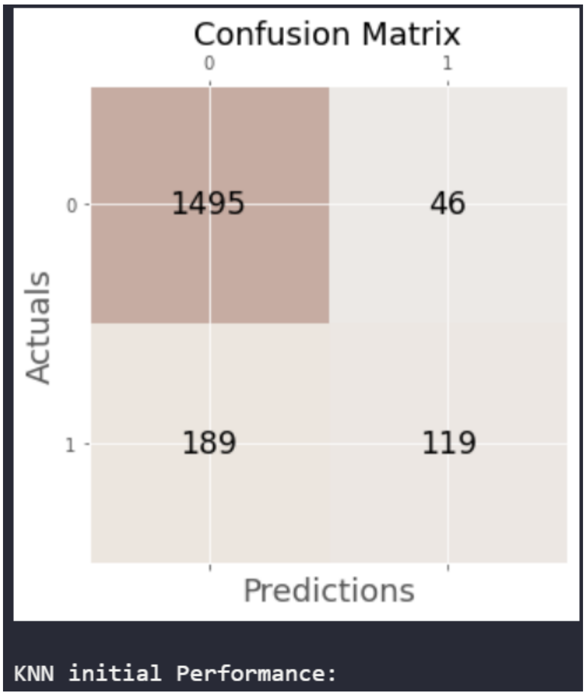


Result

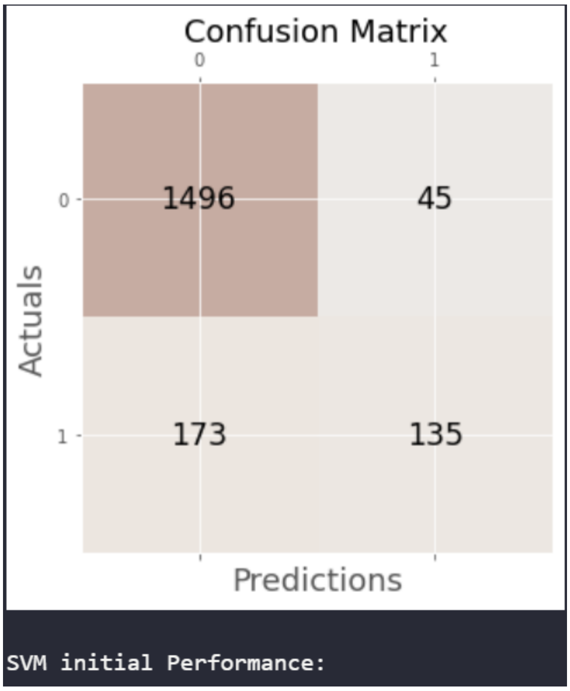
Algos	**Accuracy**	**Precision**	**Recall**	**F1 score**	**Sensitivity**	**Specificity**	**AUC**	**Error rate**	**TPR**	**FPR**	**Execution time**
Logistic regression	0.877	0.731	0.416	0.530	0.416	0.970	0.693	0.123	0.416	0.030	0.164
KNN	0.873	0.721	0.386	0.503	0.386	0.970	0.678	0.127	0.386	0.030	0.170
SVM	0.882	0.750	0.438	0.553	0.438	0.971	0.705	0.118	0.438	0.029	4.585
RBF SVM	0.888	0.758	0.477	0.586	0.477	0.970	0.723	0.112	0.477	0.030	1.752
Naïve Bayes	0.799	0.415	0.630	0.500	0.630	0.822	0.726	0.210	0.630	0.178	0.025
DT (Info gain)	0.890	0.697	0.597	0.643	0.597	0.948	0.773	0.110	0.597	0.052	0.026
DT (Ginni Index)	0.888	0.700	0.575	0.631	0.575	0.951	0.763	0.112	0.575	0.049	0.058
Random Forest	0.902	0.762	0.594	0.668	0.594	0.963	0.779	0.098	0.594	0.037	0.874
Gradient Boost	0.904	0.750	0.633	0.687	0.633	0.958	0.795	0.096	0.633	0.042	1.343
AdaBoost	0.889	0.713	0.555	0.624	0.555	0.955	0.755	0.111	0.555	0.045	0.468



Logistic Regression initial Performance:

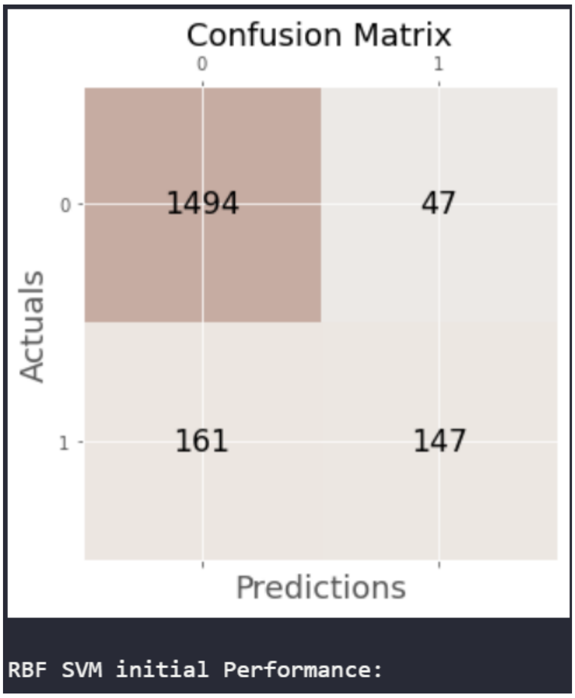


KNN initial Performance:



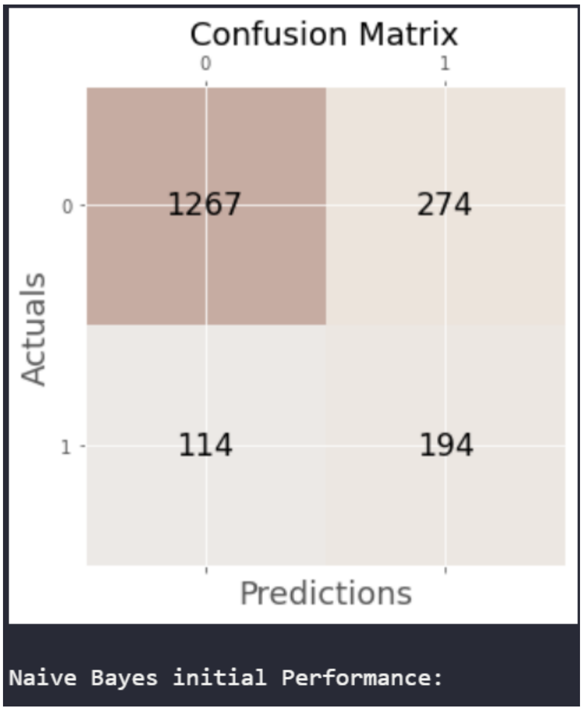
	0	1
0	1496	45
1	173	135

```
SVM initial Performance:
```



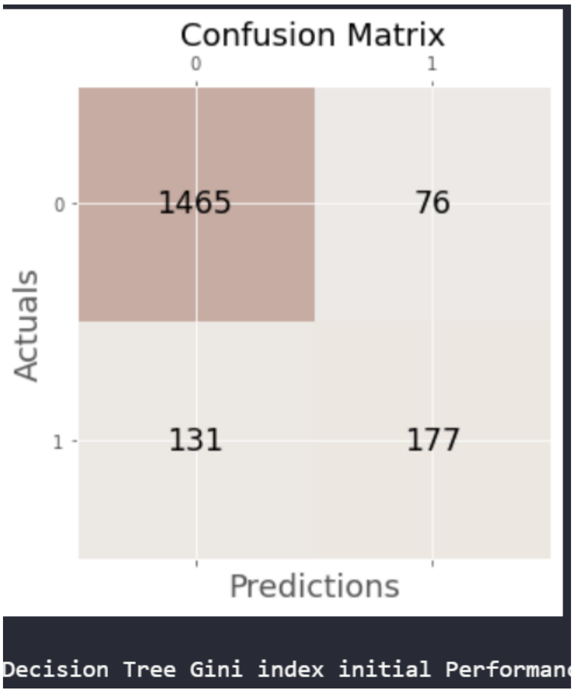
	0	1
0	1494	47
1	161	147

```
RBF SVM initial Performance:
```



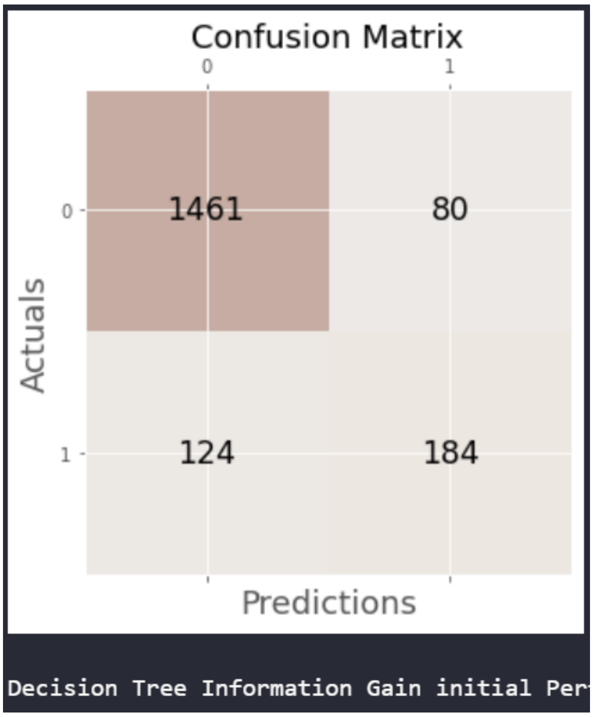
	0	1
0	1267	274
1	114	194

```
Naive Bayes initial Performance:
```

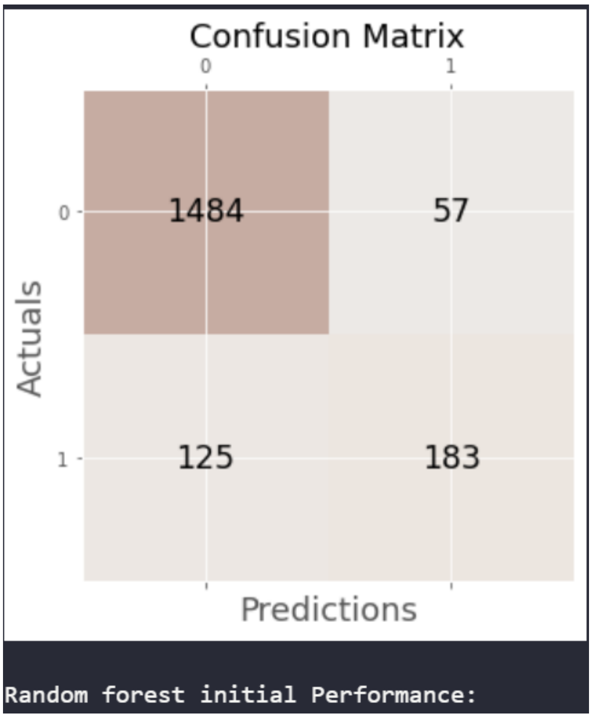


	0	1
0	1465	76
1	131	177

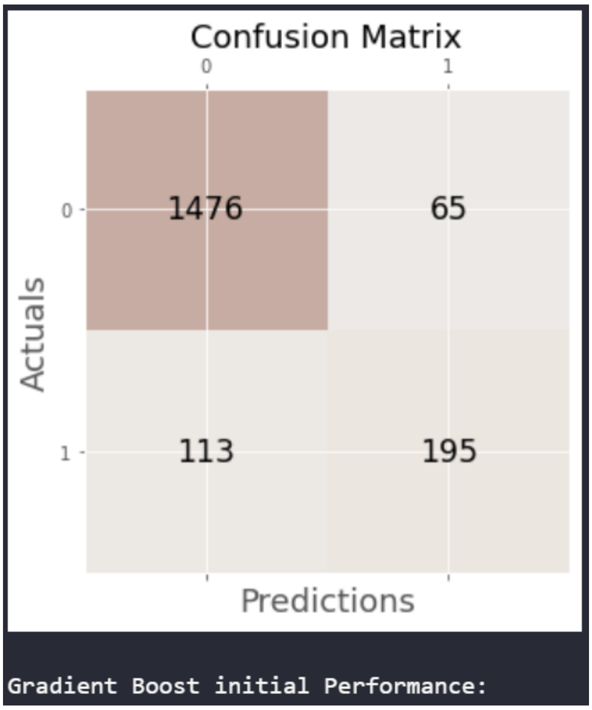
```
Decision Tree Gini index initial Performan
```



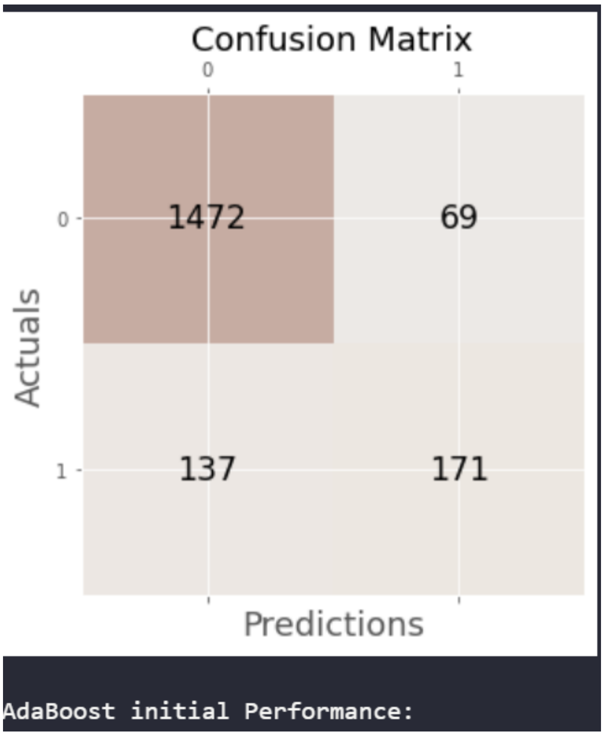
```
Decision Tree Information Gain initial Per
```



```
Random forest initial Performance:
```

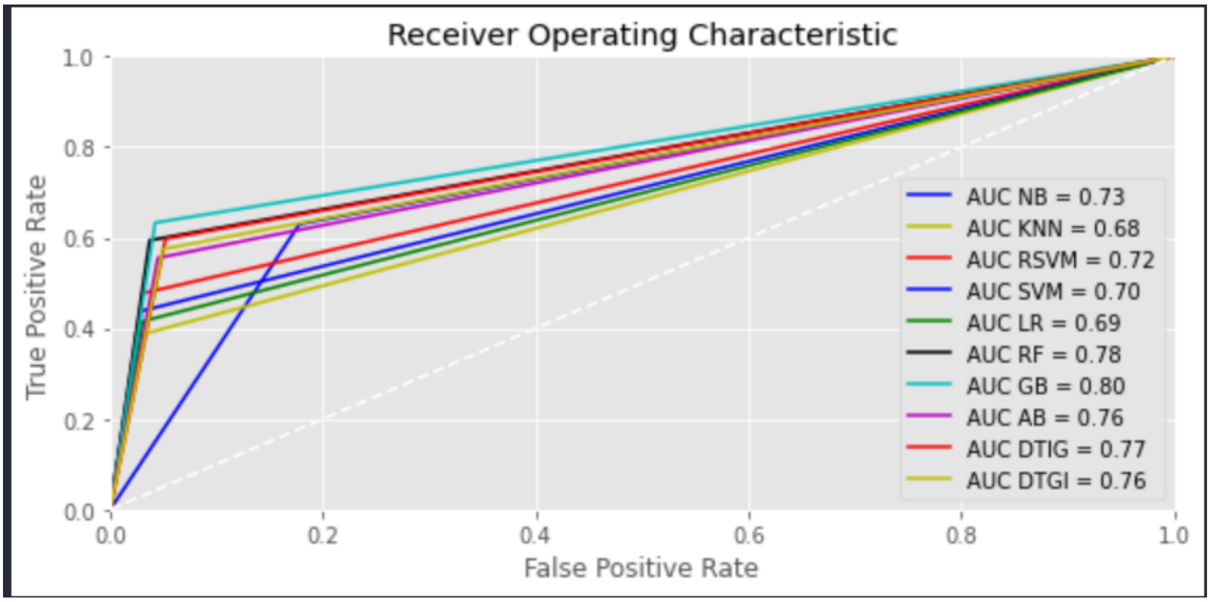


```
Gradient Boost initial Performance:
```



```
AdaBoost initial Performance:
```

ROC Curves



It is clear from all the calculated classification metrics shown above that **Gradient Boosting** classifier is the one with the highest performance, and thus, it is the one we will continue to choose.

Conclusion

In this project, we used *Online Shoppers Intention* dataset to build models that can classify website visitor, and predict which of them is likely going to make a purchase on the website. 10 different learning classifiers (Naive Bayes, KNN, SVM, Logistic Regression, Random Forest, Gradient Boosting, Decision Tree and Adaboosting) were tested and optimized, and we have achieved the best classification performance using Gradient Boost classifier, followed by random Forest, and then Adaboost.

The best classification performance:

Accuracy: 90.4%

F1 Score: 0.687

References

[1] Statistics and facts about global e-commerce. Retrieved from

https://www.statista.com/topics/871/online-shopping, 2017

[2] Statistics and facts about e-commerce in India. https://www.sta

tista.com/topics/2454/e-commerce-in-india, 2017

[3] Number of active buyers on Alibaba in Bangladesh from 2015 to 2017,

https://www.statista.com/statistics/898967/bangladesh-e-commerce export-distribution, 2018

[4] Aghdaie, Mohammad Hasan, Sarfaraz Hashemkhani Zolfani, and Ed mundas Kazimieras Zavadskas. ”Synergies of data mining and multiple

attribute decision making.” Procedia-Social and Behavioral Sciences 110

(2014): 767-776.

[5] Kumar, Anil, and Manoj Kumar Dash. ”Factor exploration and multi criteria assessment method (AHP) of multi-generational consumer in

electronic commerce.” International Journal of Business Excellence 7.2

(2014): 213-236.

[6] Rygielski, Chris, Jyun-Cheng Wang, and David C. Yen. ”Data mining

techniques for customer relationship management.” Technology in society

24.4 (2002): 483-502.

[7] Seng, Jia-Lang, and T. C. Chen. ”An analytic approach to select data

mining for business decision.” Expert Systems with Applications 37.12

(2010): 8042-8057.

[8] Huiying, Zhang, and Liang Wei. ”An intelligent algorithm of data pre processing in Web usage mining.” Fifth World Congress on Intelligent

Control and Automation (IEEE Cat. No. 04EX788). Vol. 4. IEEE, 2004.

[9] Plewczynski, Dariusz, Stphane AH Spieser, and Uwe Koch. ”Assessing

different classification methods for virtual screening.” Journal of chemical

information and modeling 46.3 (2006): 1098-1106.

[10] Benediktsson, Jon Atli, and Philip H. Swain. ”Consensus theoretic clas sification methods.” IEEE