Predicting the Outcome of Purchasing a Hotel Package to Assist Policymakers Using Machine Learning

by

Zahedul Islam
18101209
Tahsina Alam Shetu
18101174
Atulan Bhattacharjee
18101521
Jannatul Ferdous Mohima
17301009

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
BRAC University
May 2022

Declaration

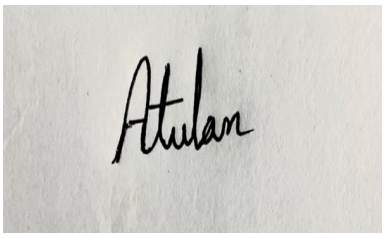
It is hereby declared that

1. The thesis we submitted was written as part of our degree program at Brac University.
2. The thesis does not contain any content that has been previously published or authored by a third party, unless it is properly cited and referenced.
3. The thesis contains no material that has been approved or submitted for any other university or other institution's degree or diploma.
4. All major sources of assistance have been acknowledged.

Student's Full Name & Signature:

Zahedul Islam
Student ID: 18101209

Tahsina Alam Shetu
Student ID: 18101174



Atulan Bhattacharjee
Student ID: 18101521

Jannatul Ferdous Mohima
Student ID: 17301009

Approval

The thesis titled "Predicting the Outcome of Purchasing a Hotel Package to Assist Policymakers Using Machine Learning" submitted by

1. Zahedul Islam (Student ID: 18101209)
2. Tahsina Alam Shetu (Student ID: 18101174)
3. Atulan Bhattacharjee (Student ID: 18101521)
4. Jannatul Ferdous Mohima (Student ID: 17301009)

The Summer of 2022 was acknowledged as sufficient in partial completion of the criteria for a B.Sc. in Computer Science on May 26, 2022.

Examining Committee:

Supervisor:
(Member)

Dr. Amitabha Chakrabarty, PhD

Associate Professor
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

Ms. Jannatun Noor Mukta

Lecturer
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Ethics Statement

This thesis is declared to be genuine and was introduced based on our research findings. All further types of information are based on these factors. The following thesis has never ever been presented, wholly or partially, to any other institution or university in order to pursue a degree.

Abstract

Given the growing and massive shock to the hotel industry, the coronavirus (COVID-19) outbreak has generated an unexpected problem. The hotel industry is currently facing a big obstacle during the COVID period, where strategies designed to promote sustainability will play an essential role for the industry. Hotel policymakers would need to learn from the COVID-19 issue in order to increase sales, improve crisis management policies, and better prepare targets and the industry as a whole to respond to unforeseen situations. It would be preferable if the hotel sector used information and technology and focused on automated ways of learning and forecasting from historical data. The objective of this research is to study customer information in order to make a recommendation to the policy maker and marketing team, as well as to develop a model to predict who will buy the newly launched vacation package. It will assist the hotel industry in enabling and establishing a viable business model for growing their customer base. As a result, we're adopting machine learning to predict which customers will be interested in purchasing the hotel package. A survey was used to gather data for this study. The data was evaluated to uncover key elements for our study, and we employed seven algorithms, namely, Random Forest Classifier, K Neighbors Classifier, Naive Bayes, AdaBoost, Support Vector Machine, Logistic Regression, and Gradient Boosting algorithm, to predict potential customers based on those features. We have obtained an accuracy of 95% while also reducing the number of false negatives by using the Gradient Boosting Classifier, Support Vector Machine, Random Forest Classifier, Naive Bayes, Logistic Regression. Our findings show that, without a team of analysts, our analysis and suggested approach can provide the finest insight to policymakers to help them make better decisions.

Keywords: Machine Learning; Prediction, Hotel Package; Policymakers

Dedication

This special thanks goes to the advisors at our university, because without them we would not have been able to complete it. Our professors have been more than the academic advisors; they have also provided guidance and support when we required it the most.

Acknowledgement

All gratitude is due to Almighty Allah, Most Kind, Most Gracious, who has provided us with the opportunity to study at BRAC University.

Without the aid of numerous people, we would not have been able to complete our bachelor's degree, let alone write this thesis. It gives us great pleasure to take this opportunity to express our gratitude for their help and advise.

Our heartfelt gratitude goes to our respected Supervisor, Dr. Amitabha Chakrabarty, Associate Professor and our Co-supervisor, Ms. Jannatun Noor Mukta, Lecturer for their invaluable assistance and support during this project. This thesis would not have been finished without their constant encouragement and guidance. We are quite fortunate to have worked with supervisors who are both motivating and approachable.

We'd also like to appreciate everyone who helped us collect data, including our friends, faculty members, coworkers, juniors, including all the responders who took time out of work schedules to take our survey. Without their assistance, this research might not have been possible.

We'd really would like to thank our loved ones, especially our parents, brothers and sisters, who are the foundation of our existence. This research would not have been possible without their unwavering love and support.

This thesis is dedicated to all of them.

Table of Contents

List of Figures

List of Tables

Nomenclature

FN False Negative

FP False Positive

KNN K-Nearest Neighbor

LR Logistic Regression

RFC Random Forest Classifier

SVM Support Vector Machine

TN True Negative

TP True Positive

Chapter 1

Overview

This chapter provides a high-level explanation of the research which was organized to address the issue, as well as the goal, objective, and technique of resolving the problem throughout this study.

1.1 Introduction

The importance of hotel businesses has been severely impacted by the COVID-19 epidemic and countermeasures are being taken to combat it. Given the growing and massive shock to the hotel industry, the coronavirus (COVID-19) outbreak has generated an unexpected problem. Currently, hotel businesses are faced with a significant challenge during Covid: sustainability. Despite growing interest in sustainability among those working in the hotel industry, there are significant differences in the frequency and quality of data and information provided by the world's leading chain hotels. Even after the fact that the overwhelming majority of the world's leading hotel chains make higher levels of commitment to sustainable development, several of them acknowledge, either explicitly or implicitly, that they are only at the beginning of what may be a long and arduous journey, and as a result, a variety of issues deserve to be discussed. In addition, it is a competitive and price-sensitive business, with the majority of customers using the Online services and a variety of browsers to determine the best discounts, offers, and substitutes with just a few clicks, as well as the hotel industry is no exception [22]. The outbreak is also affecting tourism data collecting even during crises, as regular sources of information and collection methods may not have been available (e.g., no surveys of travelers at crossings, or data provided by shuttered accommodation providers and other tourist enterprises). This has concerns for the accuracy of public tourist information once they become available, and it will necessitate estimations, maybe based on alternative sources of data. Companies are at various stages of COVID-19 crisis management, and while some are changing regulations to address gaps and the demands of hotel industries, others have been beginning to prepare full industry recovery plans. Whereas the focus has rightfully been on ensuring safety and visitors and assisting businesses in surviving the crisis, policymakers are also considering the sector's long-term implications and the fundamental changes that will be required to build a better, more sustainable, as well as adaptable tourist industry market in the long term [19]. The sustainable transition and digitalization will remain crucial in the aftermath of the immediate emergency response, and policymakers' decisions

will shape the hotel industry in the post-COVID-19 situation. In addition to these direct remedies required, policymakers would need to learn from the COVID-19 situation in order to enhance crisis management strategies and better equip targets and the industry as a whole to respond to unforeseen events. Hotel companies and tourist attractions would need to adapt their offerings to accommodate changing travel habits.

1.2 Problem Statement

The hotel industry has perhaps been slow in adopting machine learning – a branch of the broader area of Artificial Intelligence – focusing on automated techniques to learn and forecast from prior data. Is it a traditional occurrence? Perhaps. In respect of its major feature – the customized, human-facing client experience – the hotel business is one of the most traditional of all, and it has found it difficult to accept machines to replace human suggestion and action. During purchasing products and services, today's customer expects more responses from hotels, and the modern customer is no exception. Traditional hotel companies must evolve and adapt to changing customer demands.
Machine learning can help businesses analyze clients' needs, preferences, including patterns in order to provide a personalized product, alternatives, or service, in this case an entire travel arrangement. Due to the obvious large amount of basic data available, it could also be used to forecast patterns and client behavior.
For fast growth of hotel business, we need to do predictive analytics for getting an overall insight into business success, which is helpful in making the required modifications and help hotel policymakers to get a better solution in the competing hotel industry. That's why, hotel owners must constantly upgrade their goods and/or services and help to understand which segment of customers they need to target. But in the perspective of Bangladesh, there is no efficient or particular method has found or done till to the date which is a major concern of competitive hotel industry. So, for the fast growing of hotel industry, this problem needs to be solved.

1.3 Aim of Study

The primary objective of this study is to find out whether or not a customer is more inclined to acquire a newly released hotel package using machine learning and data evaluating techniques by just completing certain basic assessments that are relevant to this research. As part of this, we really would like to make it easy for the company to deploy its resources to identify as a priority who will be interested in owning the hotel package. The organization can utilize this tool to identify the most important aspects that will influence whether a customer purchases a product or not. In order to run our model and predict whether a customer will purchase a hotel package or not, we used seven different algorithms. A number of techniques are used, Random Forest Classifier, K Neighbors Classifier, Naive Bayes, AdaBoost, Support Vector Machine, Logistic Regression, and Gradient Boosting algorithm are the algorithms used to select the most important features or factors that are associated with Bangladeshi people who are interested in traveling or planning to travel

while eliminating the most irrelevant features or reasons that are not associated with it. This research was thoroughly discussed in chapters 3 and 4 of our paper.

1.4 Research Methodology

Basically, we focus our research on gaining a complete understanding of the customer's information and assisting in the development of the hotel business suing machine learning approach. The work with a strong attention on giving a representation of the state - of - the-art of scientific investigations, which usually provides a much more critical viewpoint on the concepts under discussion, impacted our technique for reviewing the literature. This study has been carried out over a long period of time. However, we will continue with this research because we intend to expand it in order to obtain more information. Following that, we began reading publications that were comparable to our concept. Our concept was still in the early stages of development at the time. At the very same time, with the help of our supervisor, we were able to successfully complete our hypothesis and begin our investigation to identify relevant explanations and aspects that are associated to this notion. We were extremely cautious when creating our questionnaire forms and we learned how to conduct a survey in a formal manner. Then we began our survey after taking all necessary precautions and formal planning. After we finished collecting data, it was important to clean and pre-process everything before applying machine learning mechanisms to it. On the 1st of January, we began drafting chapters 1, 2 and 3 of our thesis paper. Following that, we spent about a month writing chapters 4 and 5. Finally, we complete the last portion of our thesis paper writing. We spent about 14 days extensively reviewing our work and then using Microsoft Word to change our writings into the BRAC university's format. Figure 1.1 depicts all of the information.

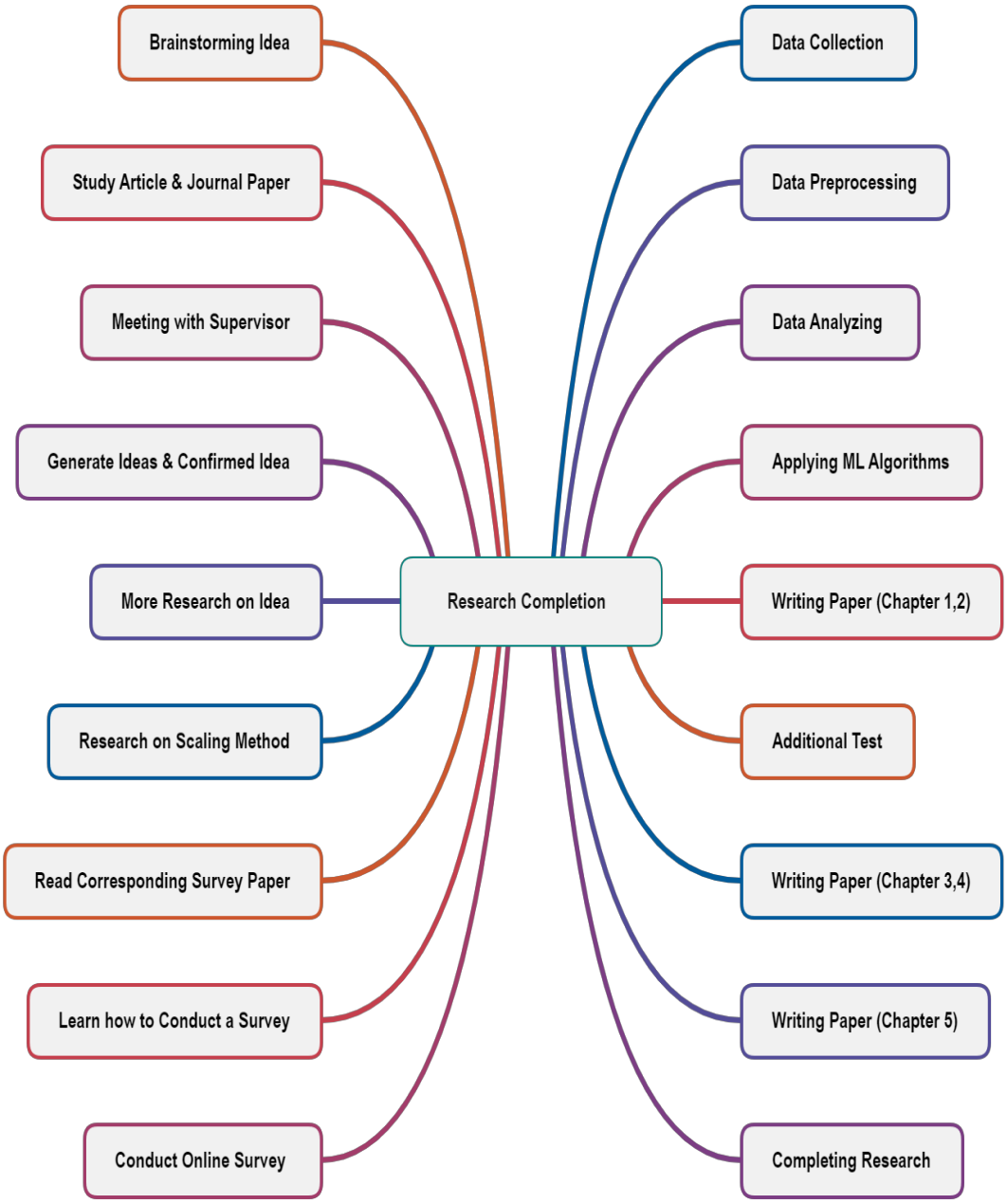


Figure 1.1: An Outline of the Research Process

1.5 Thesis Outline

- In chapter 2, background studies on machine learning algorithms and past work linked to this topic are reviewed, as well as a quick introduction to the hotel sector, inventing and engaging in the hotel industry, and how to choose the appropriate hotel package using a common scaling method.
- In chapter 3, there is a complete representation of the data collection method, an analysis of relevant factors for purchasing a hotel package, an overview of data preprocessing and reprocessing of observations, and finally an explanation of the finished visualization techniques.
- In chapter 4, it contains a thorough explanation of the exploratory analysis and findings, including the process of implementing the automated system, cross-validating the statistical sample, determining the accuracy of various machine learning algorithms, and a comprehensive review of the procedure of finding the optimum features that can ultimately lead to description.
- In chapter 5, it brings the paper to a close with general observations, shortcomings, recommendations, and suggestions for further studies relating to our research.

Chapter 2

Background Study

This chapter goes on to explain the issue in terms of our selected respondents, the modulation procedures followed, and reviews of previous research.

2.1 Introduction

When doing research, one must proceed through a number of stages in a specific order. Both the subject matter and the areas of focus for the research are selected first. After then, existing studies articles on the topic are reviewed for the purpose of obtaining background knowledge. Ideas are developed through the process of conducting background research, and as a result, this study can help establish which research activities need to be expanded upon. The next possible step is to put the findings into practice by employing a research methodology as an option. For instance, the gathering of data, the processing of data, and similar activities. In addition to this, an examination of the investigation's outcomes and results is carried out. After that, an evaluation will take place. After the conclusion of the study, there is discussion regarding the implications of the findings for the years to come.
In today's competitive business environment, knowing the different aspects that influence how every other industry performs has become essential for businesses across all sectors to explore into innovative ways of enhancing corporate performance and profitability. Predictive analytics is critical in offering a better overall insight into business success, which is helpful in making the required modifications to continue improving. To be competing in the hotel industry, businesses must constantly upgrade their goods and/or services. It differs greatly depending on the interests and background of each individual client, making it more challenging for businesses to address the identified each demand effectively.
Basically, our main research is to predict the outcome of purchasing a hotel package to help hotel policy makers to make a good decision in favor of hotel industry. The results of the data - driven models proposed in this work will assist the hotel policymakers in effectively managing their clients and offering decent customer 's needs.

2.2 Related Work

In order to identify specific customers into categories and forecast their purchase behavior, Jeyaratnam et al. [17] examines client data from a specific tourism operator and translates the data into relevant insights. In all industries, the internet offers a new way to do business. In all sectors of the economy, particularly tourism, Internet technology allows for maximum and process improvement [19]. Andres et al. [18] created an interactive and statistics system for forecasting if a client will make purchases in the coming years and proposed a different collection of consumer attributes derived from the timings and rates of past purchases for this reason. To anticipate satisfaction of customers with hospitality business, researchers used consumer hotel comments and suggestions, hotel details, and photos. Theoretical and practitioner consequences for the hospitality business is offered based on the results of this study [21]. Using 'big data' collected from online social networking sites contexts, the suggested hybrid machine learning techniques can be used as an incrementally indicates that consumer for spa hotel/resort categorization [16]. In the e-commerce arena, recommendation system has arisen, and they will be designed to proactively promote the proper things to online consumers.
Traditional Collaborative Filtering (CF) recommendation system link similarity between users by recommending goods depending on existing single-rating input. Multi-criteria evaluations are utilized rather than single-rating response in multi-criteria CF recommendation system, which can greatly enhance the accuracy of typical CF algorithms. These systems have been implemented effectively in the tourism sector. In this research [14], they used clustering, function approximation, and prediction approaches to present a new process of generating described as a multi-CF to improve the prediction performance of recommender systems in the tourism destination.
This research [20] analyzes several approaches for anticipating short-term hotel demands for lead durations of 14 days. Because of the double temporality, machine learning techniques are evaluated to methods ranging from seasonally naive to simple exponential methodologies. A new strategy based on resolving disputes, in which numerous estimation techniques are dynamically integrated to get predictions, is among the machine learning methods studied. In the hotel industry, reservation cancelation has a substantial impact on demand managerial decisions. Hotels adopt strict reservation rules and online booking strategies to minimize the impact of cancellations, which could have a detrimental effect on income and the hotel's prestige. A machine learning-based prototype of the system was created to mitigate this consequence [12].

2.3 Machine Learning

2.3.1 What is Machine Learning (ML), and how does it work?

ML is a subfield of AI that permits mechanisms to gain an understanding of and develop without ever being explicitly programmed. ML is concerned with the creation of computer software that can examine information and data for itself. Our ability to consistently implement ML in the world depicted. Nevertheless, turning machines

into intelligent robots is not as simple as it appears. ML is required for powerful AI to help machines comprehend like humans can. ML, like the human psyche, depends on inputs, including such testing phases or visualization techniques, to grasp entities, subdomains, and their interconnections. ML can begin once objects have been defined. Assumptions or information, including such instances, actual experience, or guidance, are used to start the machine process of learning. It searches for patterns or relationships so that it can draw conclusions based upon the information presented. The basic goal of machine learning is to allow machines to understand without human involvement and change their operations accordingly. In a machine learning model, train data represents experiences and is fed into the algorithms that learn and train them, with both the outcome resembling experience or skills obtained from the data. It is to find the best feasible sets of functional h: XY [8] scientifically.
ML has been around for about some period as a hypothesis. Arthur Samuel, a computer scientist at IBM as well as a pioneer in AI and computer gaming, created the term "machine learning." Samuel created a checkers-playing software program. The longer the program was used, the more someone learned from its mistakes and reached a conclusion employing algorithms. Machine learning is a field that studies the research and development of systems that can learn from and predicts the outcome variable. The value of machine learning is that it can solve issues at a remarkable speed that human thought cannot match. Machines can be programmed to discover patterns in and correlations between incoming data and automating regular activities using huge amounts of computer power behind a specific task or numerous specific activities. Machine learning is not really a sci-fi concept. Businesses from various sectors were already using it to boost innovation and enhance operational efficiency.

2.3.2 Supervised Learning

To anticipate future ones, supervised machine learning algorithms apply what has already been learnt in the past to fresh data using annotated data. The learning algorithm creates an inference function to anticipate correct output through evaluating a predefined training set. After enough training, the system can also provide recommendations for any input vector. It could also contrast its outputs to the accurate, intended outcome to identify faults and make necessary model modifications. Our study is mainly based on supervised learning. Since we have both input as well as output, this is the case [9]. This learning data consisted of multiple input and output pairs. S = ((x1, y1)...(xm, ym)), wherein S is just the training data, which contains experimental input and output for the models to be trained. Labels are recognized in supervised learning. Here, Fig. 2.1 is the workflow of it. Detecting spam mail as well as distinguishing it from non-spam mail. These algorithms will learn which types of spam mail are spam and which are not from a training dataset of spam and non-spam mail. It can then use this training set to create a model that can distinguish between spam and non-spam messages in the future.
The creation of data driven programs to just provide genuine insight into a variety of business pieces of information is a frequently used application for supervised learning approaches. This allows companies to forecast intended results based on a particular output factor, supporting managers in defending actions or turning for the benefit

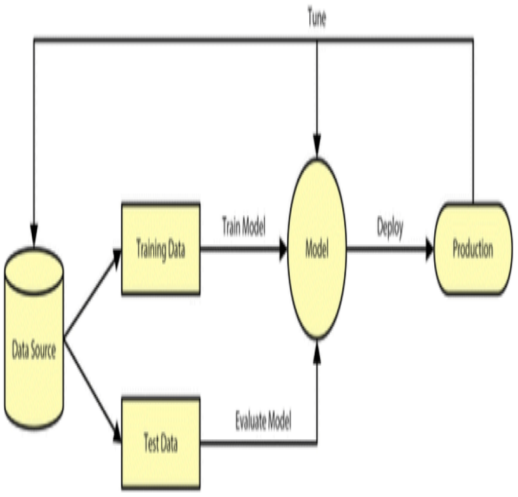


Figure 2.1: Work Flow of Supervised Machine Learning

of the company. Supervised learning models could be a useful tool for automating classifying as well as making predictions using labeled training data. To prevent overfitting data models, you'll need to structure your ml algorithms with human skill and experience.

2.3.3 Algorithm Used

Considering our present dataset, we employ multiple supervised ML algorithms to discover the best one of these for our systems. We used seven separate classifiers to train our dataset to determine the one which made the greatest results. Random Forest Classifier, K Neighbors Classifier, Naive Bayes, AdaBoost, Support Vector Machine, Logistic Regression, and Gradient Boosting are the algorithms used. A basic description of the algorithm has been provided:

- Random Forest Classifier

  Random forest is a data classification and prediction supervised learning method. However, it is most commonly employed to solve categorization issues. This is based on ensemble methods, which is a method of integrating numerous classifications to solve complicated problems and increase the effectiveness of the algorithm [10]. Instead of depending on a single tree structure, the random forest considers the forecasts from every tree as well as anticipates the final outcome depending on the majority votes of predictions. The more trees in the forest, the more accurate it is and the concern of overfitting is avoided. As a result, basic assumptions for an improved Random Forest classifier are as follows: 1. The dataset's featured variables should have some real values such that the classification can forecast reliable data instead of guesses. 2. Each tree's predictions must have some really weak correlation. The Random Forest algorithm is demonstrated in Fig. 2.2 below:

  Hyper - parameters, which have been created in random forest algorithms, are widely used to enhance the model's efficiency or forecast accuracy. Overfitting is avoided by using a random generator. In a random forest, training a model

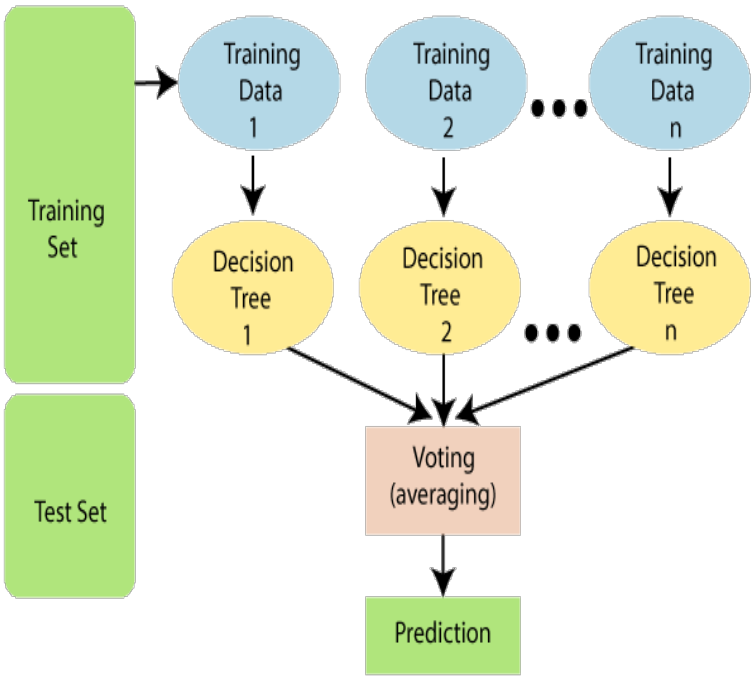


Figure 2.2: Work Flow of Random Forest Classifier

is really relatively simple as well as rapid; nevertheless, predicting the ultimate outcome in real time would take a long time. Additional benefit of random forest would be that it can be used to retrieve the crucial elements that are required for prediction while omitting all of the non-essential features, resulting in improved accuracy and quicker and more efficient prediction.

- K Nearest Neighbors Classifier

  K-NN has been one of ML's relatively basic but crucial categorization algorithms. Recognition system, information extraction, and intrusion prevention are just a few of the applications it uncovers in the supervised learning arena. The K-NN method assumes that perhaps the incoming case/data and existing instances are comparable and places the new case in the categories that are most similar to the existing category [13], [1], [5]. The K-NN method maintains all of the existing evidence and classifies a new information point depending on its similarity to the existing data. This implies that small data can be quickly sorted into a suitable category through using the K-NN method [11]. The K-NN algorithm could be used for both classification and regression problems, but it is more commonly utilized for classification techniques. It's also known as a passive learner algorithm since it doesn't learn from either the training sample right away; instead, it saves the dataset and then takes the appropriate action on it during classifications [15]. The KNN method simply saves the dataset during the training cycle, because when it receives new information, it classifies it into a category that is quite close to actual data. KNN Classifier algorithm is demonstrated in Fig. 2.3 below:

- Naive Bayes

  The Naive Bayes method is a supervised learning method for addressing classification issues that is dependent on the Bayes theorem. It is most commonly

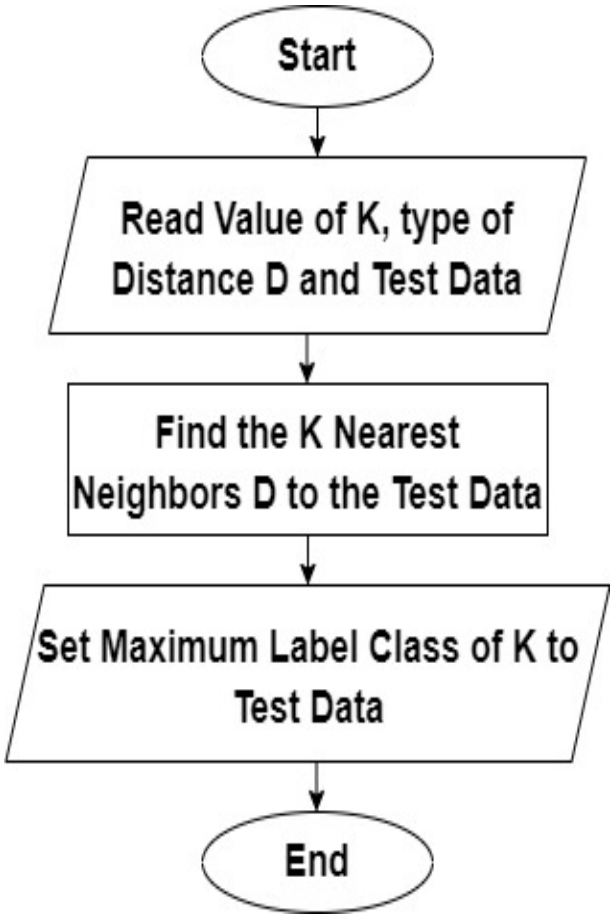


Figure 2.3: Flowchart of K Nearest Neighbors Classifier

employed in text categorization with a large - scale training dataset. The Naive Bayes Classifier [4] is a simple and efficient classification technique that aids in the development of fast machine learning methods capable of making speedy predictions.

It works for both binary and multi-class classifications. In comparison to the other Algorithms, it shows better performance in multi-class prediction. Junk mail filtration, sentiment classification, including article classification are all implementations of the Naive Bayes Algorithm. Naive Bayes algorithm is demonstrated in Fig. 2.4 below:

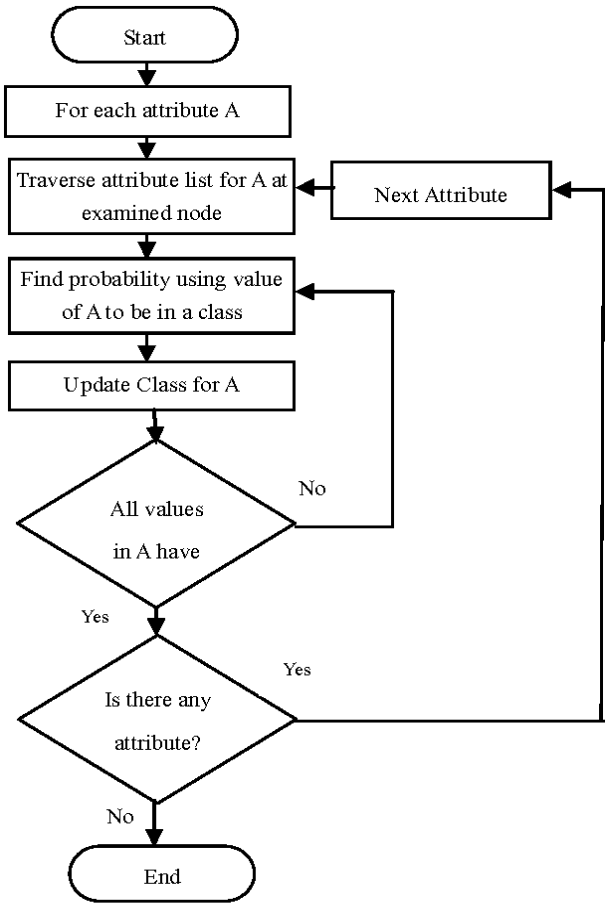


Figure 2.4: Flowchart of Naive Bayes

- Logistic Regression

  The method of predicting the possibility of a distinct result representation of the input parameter is known as logistic regression [2]. The most frequent logistic regression models have a binary outcome, which might be true or false, yes or no, and so forth. So rather than constructing a regression model, we fitted a "S" formed logistic function that indicates two peak values in regression models (0 or 1).

  It can generate chances and categorize new data using both continuous or discrete datasets, regression analysis is a key machine learning algorithm. Logistic regression could be used to categorize observations based on many types of

information and can quickly identify the most useful factors for classification. Logistic Regression algorithm is demonstrated in Fig. 2.5 below:

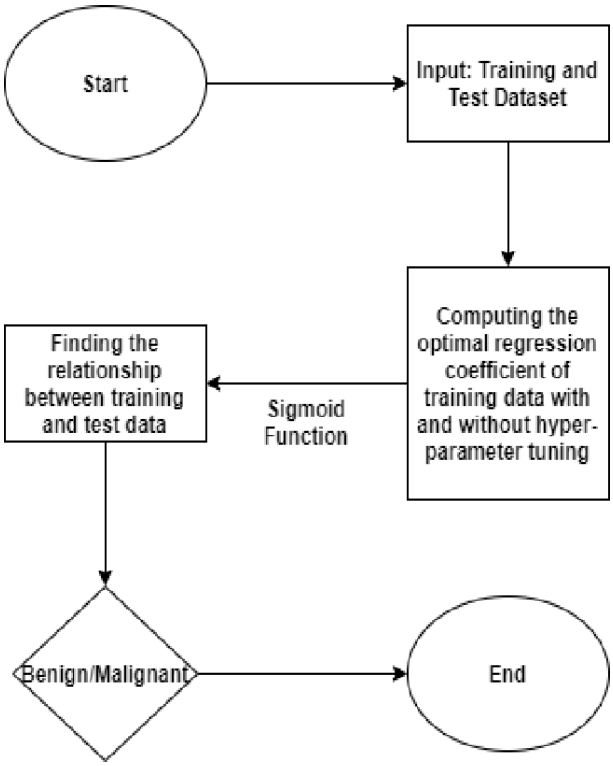


Figure 2.5: Flowchart of Logistic Regression

- Support Vector Machine

  SVM is among the extensively employed Supervised Learning algorithms for Regression and Classification issues. Nevertheless, it is commonly used in ML for Classification problems. Its goal is to find the best lines or target variable for categorizing n-dimensional spaces into subclasses so that additional data points can be readily placed in the appropriate category in the hereafter. A hyper - plane is the name for the optimal choice boundaries. The maximum arguments that assist in creating the hyper - plane are chosen via it. Support vectors are the ultimate instances, and the technique is called a SVM [3]. It works well in scenarios with a lot of dimensions. It keeps data by using a subgroup of training examples termed coordinates in the decision boundary. For the selection functions, several kernel functions can be given, as well as bespoke kernels. SVM algorithm is demonstrated in Fig. 2.6 below:

- Gradient Boosting Classifier

  A prominent boosted algorithm is gradient boosting. Every indicator in gradient boosting rectifies the error of its predecessor. Unlike AdaBoost, the trained instance parameters are not adjusted; instead, each prediction is given training that uses the immediate predecessor residual errors as labeled. CART (Classification and Regression Trees) is the fundamental learner in a technology known as Gradient Boosted Trees [6]. A total of N trees comprises the ensemble. The features matrices X and labels y are used to train Tree1. The remaining errors r1 in the training dataset are calculated using the y1(hat)

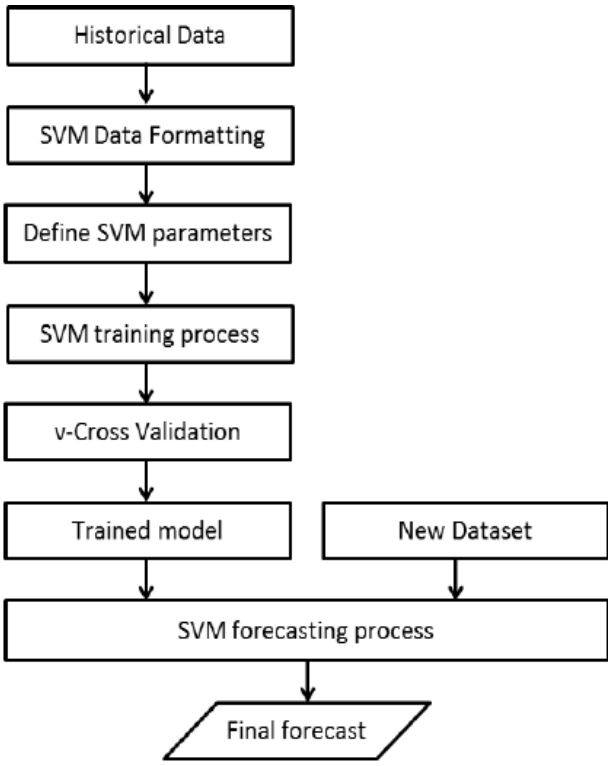


Figure 2.6: Flowchart of Support Vector Machine

predictions. The feature set X and the remaining errors r1 from Tree1 are then used as labels to train Tree2. After that, the residue r2 is calculated using the projected findings r1(hat). This procedure is done until all of the ensemble's N trees have been trained. The Fig. 2.7 illustrates how prediction problems have been solved using gradient boosted trees:

- AdaBoost Classifier

  Ensemble machine learning methods include AdaBoost models. We can quickly get a much clearer sense of how the whole model would work by looking at the actual meaning of the word 'ensemble.' Ensemble models assume the responsibility of merging multiple models to build a more advanced/accurate meta model. In comparison to its similar equivalents, this meta model seems to have a relatively good accuracy in terms of prediction. The AdaBoost [7] method is classified as an ensemble boosting strategy since it integrates new variations to create more accurate products in different stages: 1. On the training set, weak classifiers learners are allowed to learn. 2. This meta-model is produced by combining various models and attempting to fix the faults made by different poor learners. AdaBoost algorithm is demonstrated in Fig. 2.8 below:

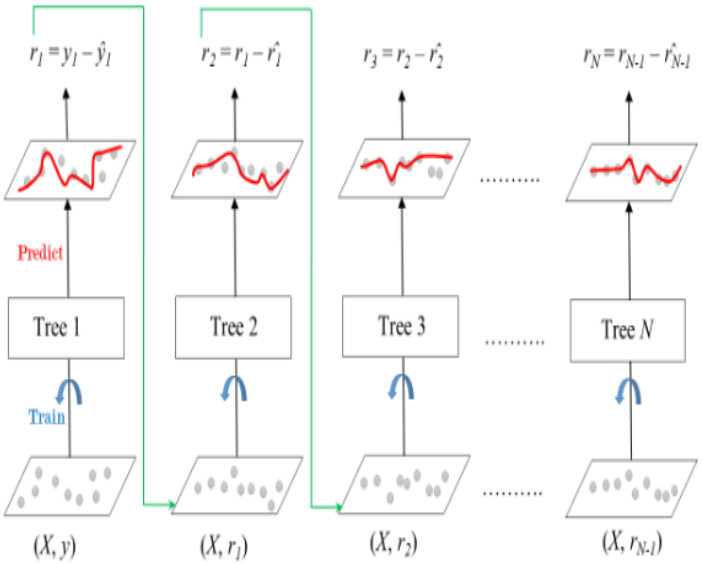


Figure 2.7: Workflow of Gradient Boosting Algorithm

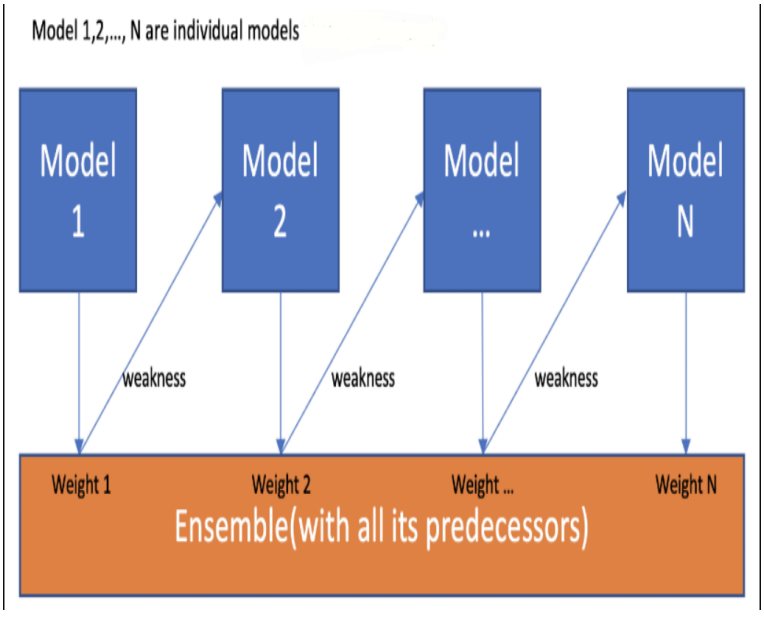


Figure 2.8: Workflow of AdaBoost Algorithm

Chapter 3

Methodology

Throughout this section, a full explanation of the process of developing the survey form, as well as a strategy for the collecting and analysis of the information, as well as findings of the data, has been offered.

3.1 Recruitment and Procedure

For the purpose of this investigation, we have compiled all of the findings from a diverse range of sources, which were provided by men and women from all around Bangladesh who took part in the questionnaire from eight distinct divisions. The vast majority of the data were obtained through internet travel and hotel groups; the remaining data were obtained from the respondents' relatives as well as their coworkers. We have collected about 2600 or more data regarding the pandemic situation with Covid-19, of which 2550 data have been included in our research. The vast majority of the information was obtained through the use of an online questionnaire, in which more than 2550 individuals took part. This questionnaire included 16 criteria that are utilized in the prediction process.
Before and during the research, protecting the confidentiality of the participants' personal information was our first priority. As a result of this, we have assured that none of the participants' personally identifying details, such as their name, address, email, or contact numbers, will be gathered from them throughout the survey. We are unable to trace or identify any user based on the data that we have collected because there is no way for us to do so. By doing things in this manner, we have ensured the protection and privacy of the data collected from our survey participants. Responses from more than 2550 persons were obtained for the survey that we conducted. In addition, we are carrying on with the collection of more data for the purpose of our ongoing research. After reviewing a large number of survey examples taken from their earlier work, we have designed our very own survey questions, making sure to adhere to the established structure and criteria.

3.1.1 Proposed Methodology Recruitment and Procedure

Under this subsection, we will outline our research methodology and processes in order to achieve our goal. Here is the approach we recommend (see Fig. 3.1). To do so, we used a variety of techniques. It all started with the data collected through an online survey.

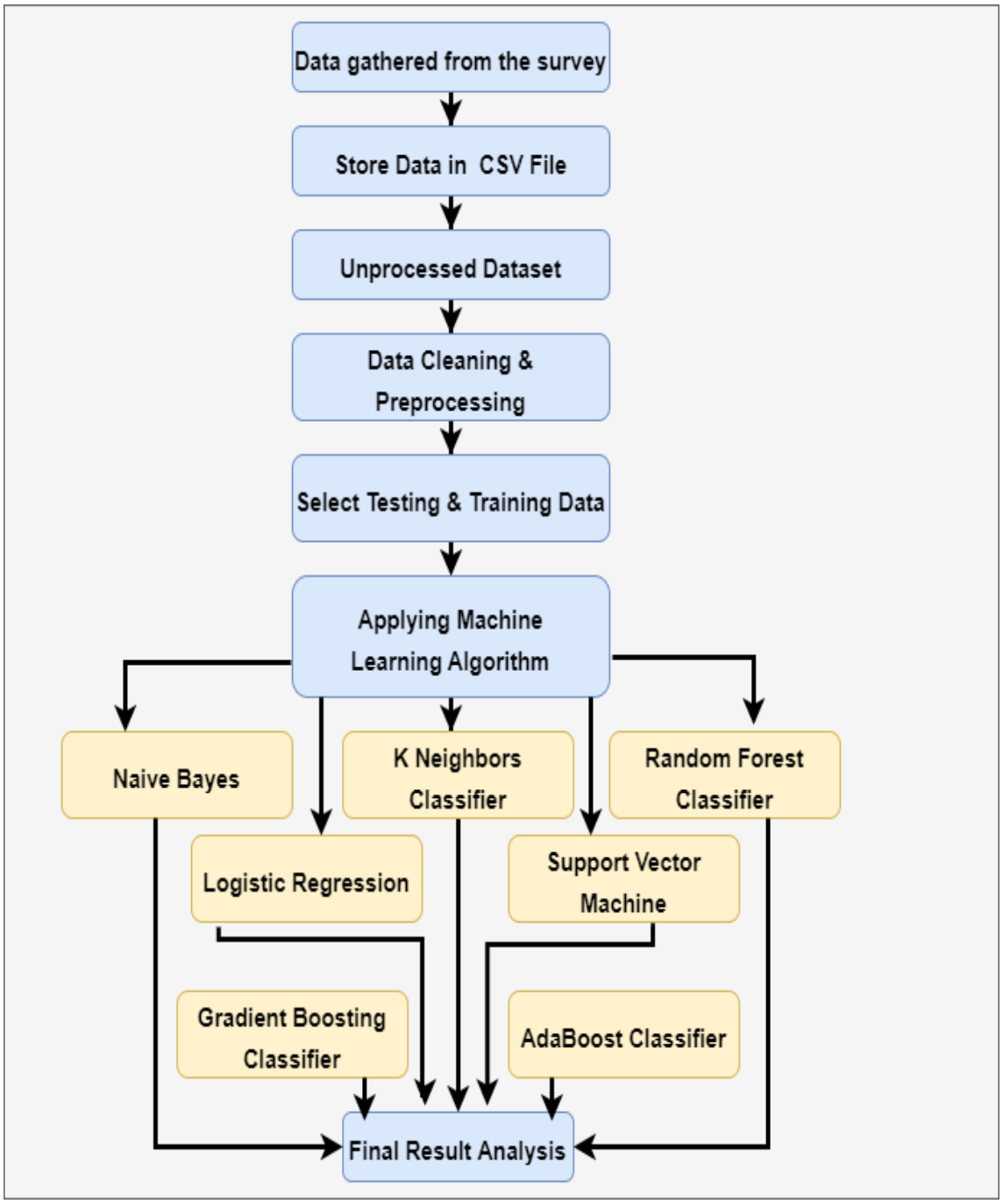


Figure 3.1: Proposed Model Flowchart

3.2 Data Set Description

Following an investigation and analysis of previous cases involving the prediction of package affordability, the study's contributory components were identified as follows (see Fig. 3.2):
'Customer ID', 'Gender', 'Age', 'Marital Status', 'Occupation', 'Monthly Income', 'Passport', 'Have Own Car', 'Purpose Of Travelling', 'Preferred Destinations', 'Total Num. of Person Visiting', 'Frequency of Travel in a Year', 'Money Spent on Vacation', 'Preferred Property Star', 'Booking Method', 'Purchased Package'

```
[8]  1  dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2550 entries, 0 to 2549
Data columns (total 16 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Customer ID                 2550 non-null   int64
 1   Gender                      2550 non-null   object
 2   Age                         2502 non-null   float64
 3   Marital Status              2550 non-null   object
 4   Occupation                  2003 non-null   object
 5   Monthly Income              2502 non-null   float64
 6   Passport                    2550 non-null   int64
 7   Have Own Car                2550 non-null   int64
 8   Purpose Of Travelling       2550 non-null   object
 9   Preferred Destinations      2522 non-null   object
 10  Total Num. of Person Visiting  2550 non-null   int64
 11  Frequency of Travel in a Year  2550 non-null   int64
 12  Money Spent on Vacation     2513 non-null   float64
 13  Preferred Property Star     2550 non-null   int64
 14  Booking Method              2549 non-null   object
 15  Purchased Package           2550 non-null   int64
dtypes: float64(3), int64(7), object(6)
memory usage: 318.9+ KB
```

Figure 3.2: Dataset Variable with Data Type

3.3 Data Cleaning and Preprocessing

Eventually, after gathering information from 2550 persons, we made the decision to begin making our prediction based on the information we had gathered. But before we put the prediction algorithms to work on the data, we cleaned the data to get rid of as many anomalies and outliers as we could. This allowed us to achieve a higher level of accuracy.
That would be the last data set toward which our seven prediction algorithms were employed. We evaluated all of the algorithms based on the most recent label and our 16 features to evaluate how accurate their predictions were. Out of the 16 features utilized for predictions, we also found out the significant features.

3.3.1 Filling Null Value

Within almost all of the 2550 records that comprise up the data set, there are some records that are missing values. In order to complete Fig. 3.3, we utilized an imputation method that included both the mean and the mode. It is a method in which the average of the obtainable cases has been used to fill in for a missing value for a particular trait. The approach is known as mean filling. This technique is straightforward to implement and guarantees an adequate number of samples.

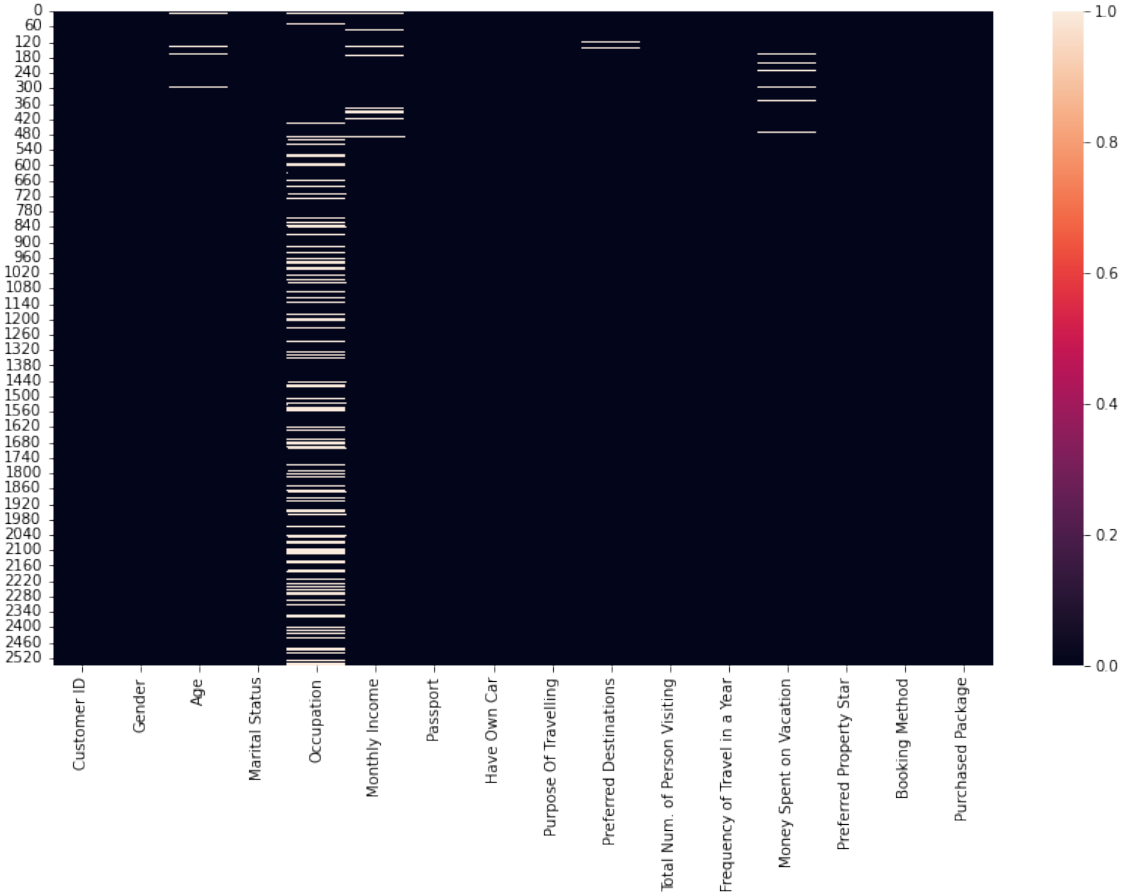


Figure 3.3: Representation of Missing Data

Once the majority of the blank field values have been filled, the anticipated prepared dataset is presented to us (see Fig. 3.4). If we examine our data set for missing values right now, we won't be finding any null values because none of them are there.

3.3.2 Data Encoding

There are a lot of different machine learning techniques, and many of them have been unable to interact directly with categorical variables. It is necessary to assign numerical values to the categories. The One Hot Encoding method is the one that we employ for this particular encoding. The process of transforming categorical variables into numerical values and then using those numerical values to integrate them into a machine learning model is an example of one hot encoding. The category variables are represented as binary vectors by the use of this encoding method. It is necessary for us to make use of dummy variables. When constructing dummy

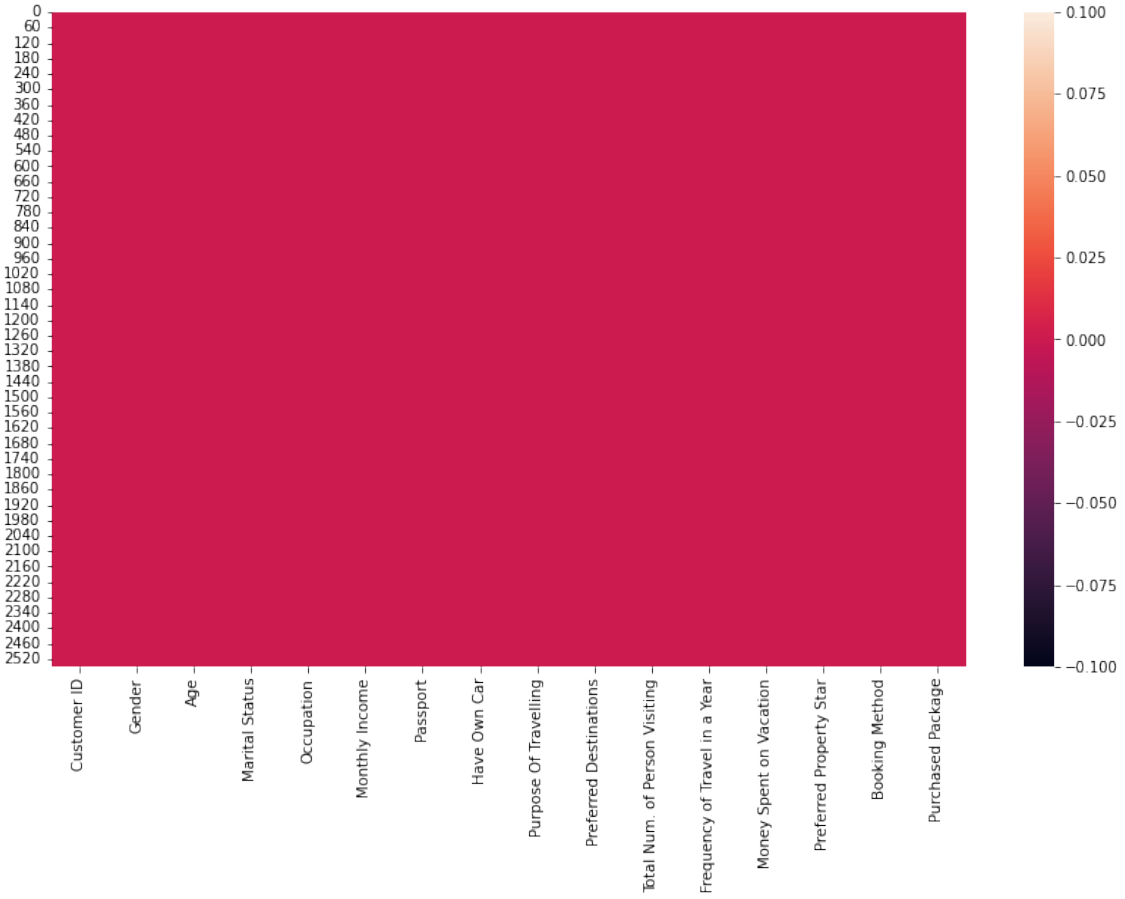


Figure 3.4: Representation of Cleaned Data

variables, one hot encoding is utilized, and these variables can only take on the values 0 or 1.

3.3.3 Feature Scaling

Considering our dataset contains some attributes with larger values, it is necessary for us to scale them out and in order to ensure that our machine learning algorithms work well. There are numerous algorithms for machine learning that call for feature scaling to be performed. It is carried out for the purpose of dealing with highly variable orders of values, magnitude or units.

3.4 Data Visualization

According to Fig. 3.5, 53% people more likely want to make a purchase of hotel package where rest of them do not want make a purchase. So, we need to find out the key factors and do more analysis to see how we can increase the rate. We can see (Fig. 3.6) that 56.1% male participants are involved in travelling, whereas female participants are less interested in travelling. Fig. 3.7, we can clearly see that most people want to travel in Bangladesh. That's why we need to give more focus on the BD hotel industry and how we can make more improvements. Fig. 3.8, we can clearly see that men have passports compared to women, and they travel the most outside of Bangladesh. That means this segment of customers we need to target more. Fig. 3.9, we can clearly see that most people want to prefer travel websites for booking hotels.

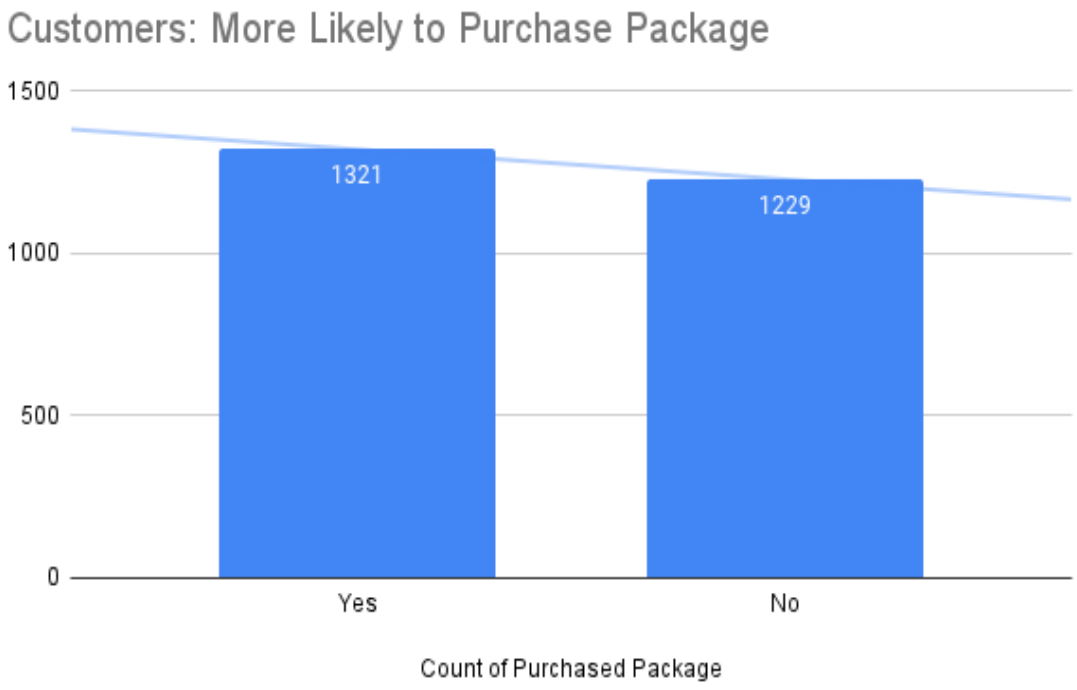


Figure 3.5: Customers: More Likely to Purchase Package

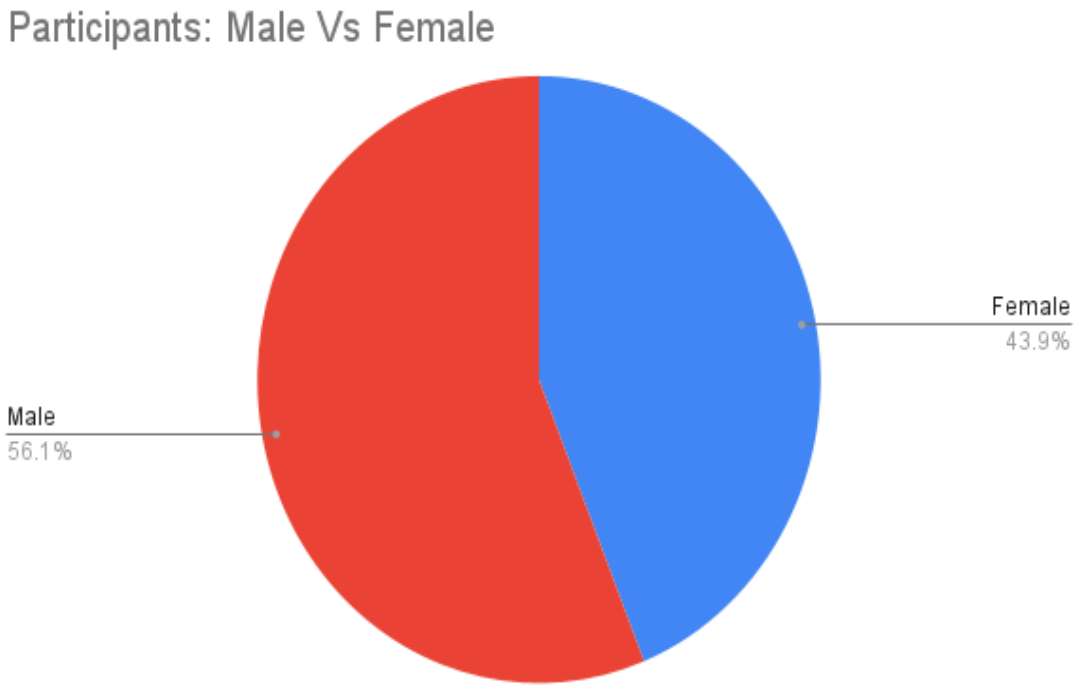


Figure 3.6: Participants Male vs Female

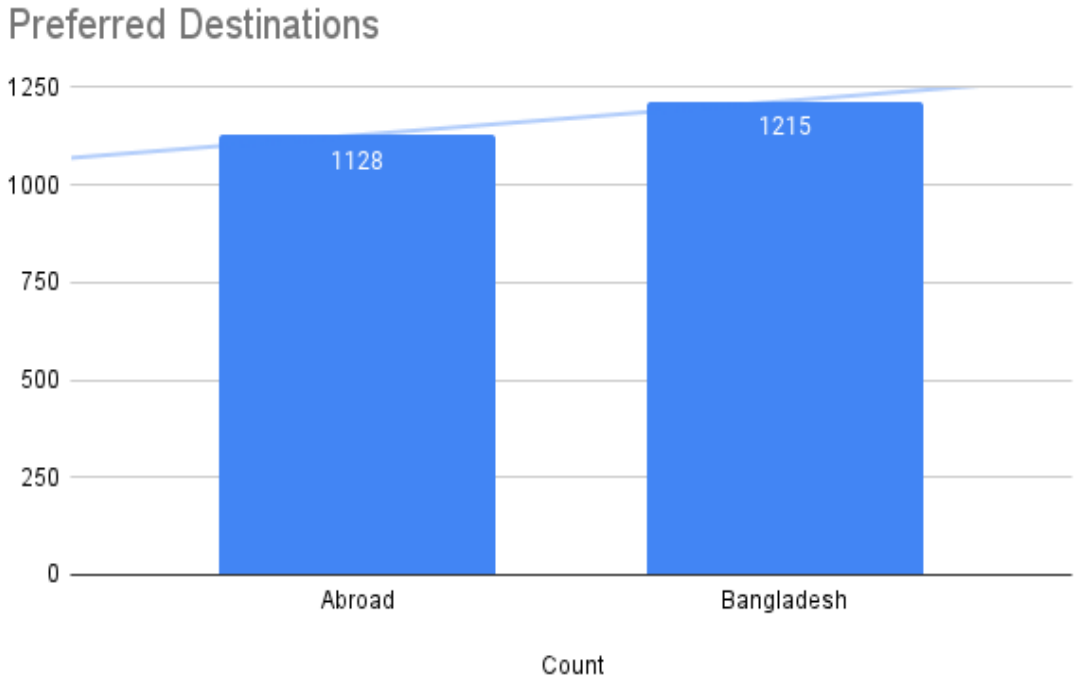


Figure 3.7: Graph of Preferred Destinations

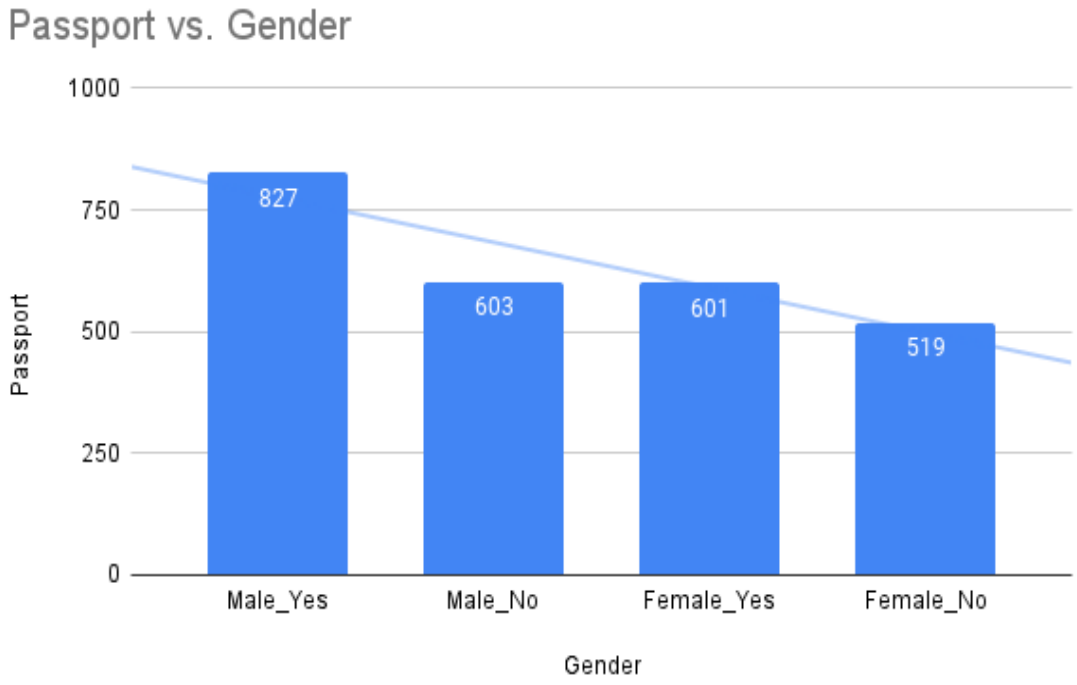


Figure 3.8: Graph Who have Passports

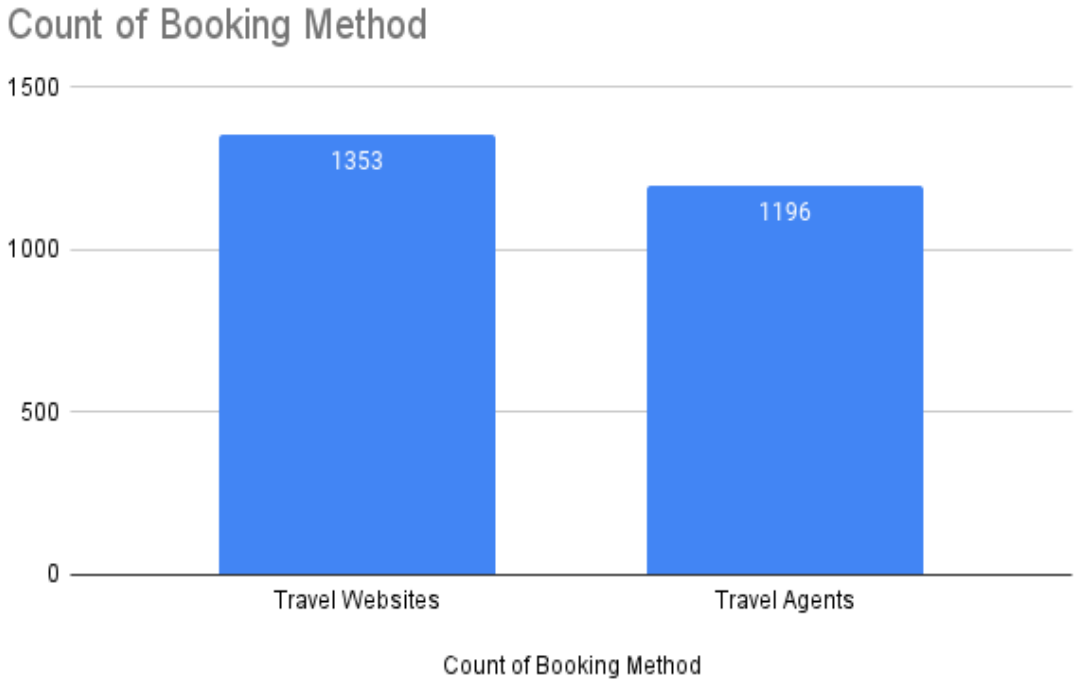


Figure 3.9: Graph of Preferred Booking Method

Chapter 4

Experimental Result and Analysis

We covered our proposed approach in the preceding chapters. We discussed how we gathered data for our system, cleaned it, and used histograms to show the final result. Following that, we employed various algorithms to identify which customer is more likely to purchase the package. The results of several algorithms will be discussed shortly. To determine the model's correctness, we had been using our 16 general features that we gathered from various articles and ran multiple machine learning techniques.

4.1 Applying Algorithms

Our research tries to determine whether or not which customer is more likely to purchase the package. Because our data set is primarily focused on binary classification, any of the two possible outcomes is possible. Our data was also visualized using histograms. We employed a couple of the most generally used methods for both of the cases described because our data is binary and we had a few incomplete data for many of the questions. To determine the accuracy, precision, recall, specificity, and f-measure of our predictive model, we used seven methods. The greater the precision and f-measure, the finer the systems will be. Precision, defined as the total number of true positives in all yes predictions, is also a significant factor. This indicates that a customer who has truly purchased the packages from all of the expected package's results in our algorithm. Remember another crucial component of the forecasting system: the number of genuine positives in genuine yes results. This signifies a customer who has truly purchased the packages from all of the expected packages in our system. Next, we'll verify the effectiveness and f-measure of each algorithm in order to determine the one which perfectly suited our model. False Negative and False Positive are two additional significant words that are critical to our system. The model will indeed be improved if the False Negative and False Positive are lower. A model would also be considered acceptable when both of them are lower and the balance is perfect.

4.2 Performance Based on Accuracy, Recall, Precision, and F1-score of Different Algorithms

Seven different classification techniques were utilized during the training of our dataset. In order to analyze categorization algorithms, we decided to use the metrics that are the most well enough and widely used. On our models, we have even employed accuracy matrices, precision matrices, recall matrices, and the F1 score to assess and examine the effectiveness of machine learning techniques. The first thing we did was divide our data into two distinct categories. When we were training our model, we utilized eighty percent of our data, and when we were evaluating the model after it had been trained, we utilized twenty percent of our data. In order to train with each of the seven different approaches, we utilized a method called test train split. See Table 4.1 Accuracy Score of Different Algorithms.

Algorithm	Accuracy
Random Forest Classifier	95.29 %
K Nearest Neighbors Classifier	83.29 %
Naive Bayes	95.29 %
Logistic Regression	95.29 %
Support Vector Machine	95.29 %
Gradient Boosting Classifier	95.29 %
AdaBoost Classifier	94.90 %

Table 4.1: Accuracy Score of Different Algorithms

4.2.1 Random Forest Classifier

Confusion Matrix:

As can be seen in Fig. 4.1, the True Negative (TN) value is 216, and the False Positive (FP) value is 21. The False Negative (FN) for the second row is set at 3, while the True Positive (TP) is set at 270.

Classification Report of RFC:

We are able to observe the values for our recall, precision, and f1-score in Table 4.2.

	Precision	Recall	F1-Score
0	0.99	0.91	0.95
1	0.93	0.99	0.96

Table 4.2: Classification Report of Random Forest Classifier

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 99 percent.

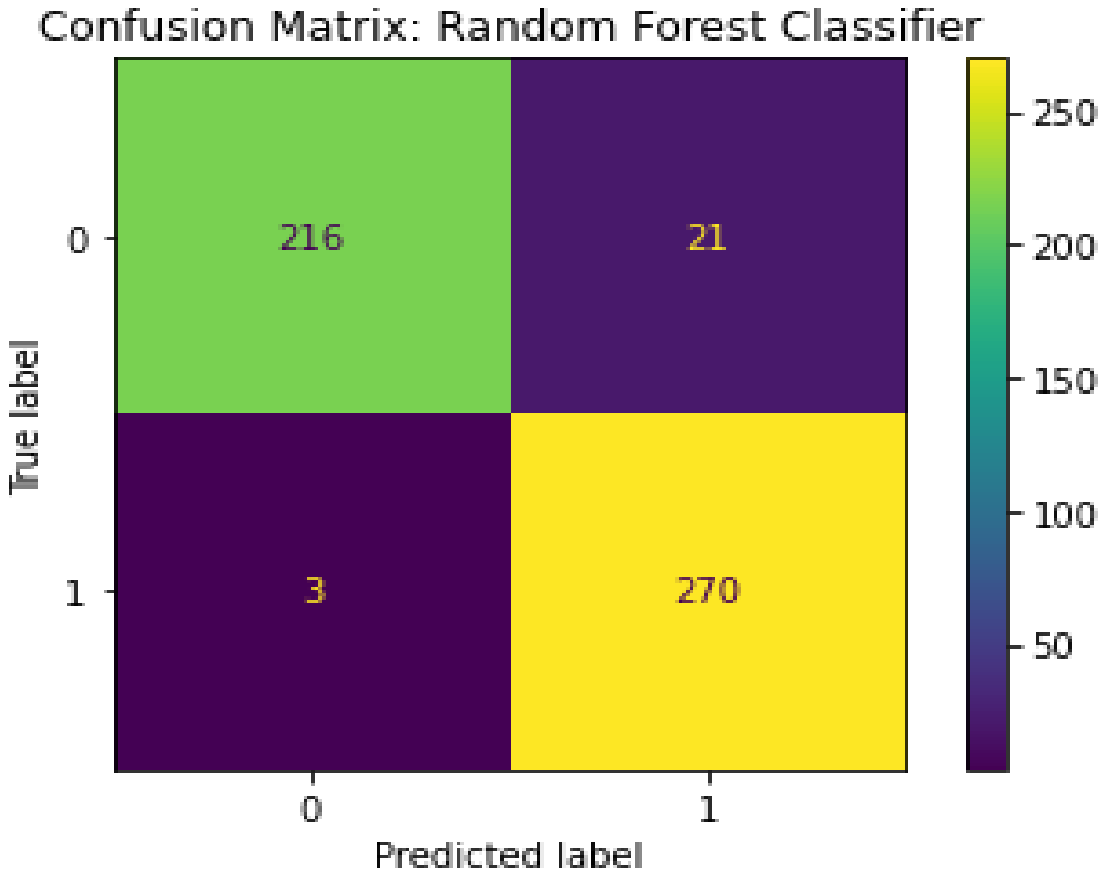


Figure 4.1: Confusion Matrix of Random Forest Classifier

After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 96 percent.

4.2.2 K Nearest Neighbors Classifier

Confusion Matrix:

As can be seen in Fig. 4.2, the True Negative (TN) value is 189, and the False Positive (FP) value is 48. The False Negative (FN) for the second row is set at 34, while the True Positive (TP) is set at 239.

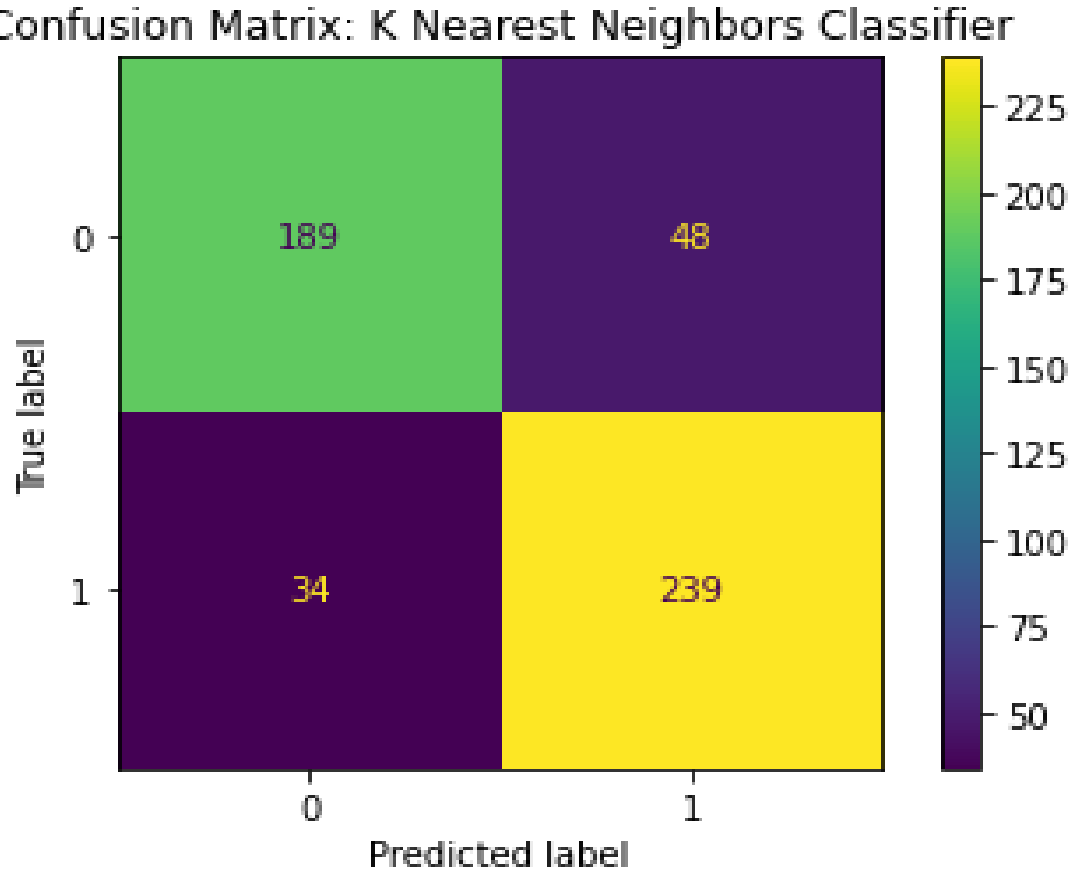


Figure 4.2: Confusion Matrix of KNN Classifier

Classification Report of KNN:

We are able to observe the values for our recall, precision, and f1-score in Table 4.3.

	Precision	Recall	F1-Score
0	0.85	0.80	0.82
1	0.83	0.88	0.85

Table 4.3: Classification Report of KNN

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 83 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 88 percent.
After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 85 percent.

4.2.3 Naive Bayes

Confusion Matrix:

As can be seen in Fig. 4.3, the True Negative (TN) value is 216, and the False Positive (FP) value is 21. The False Negative (FN) for the second row is set at 3, while the True Positive (TP) is set at 270.

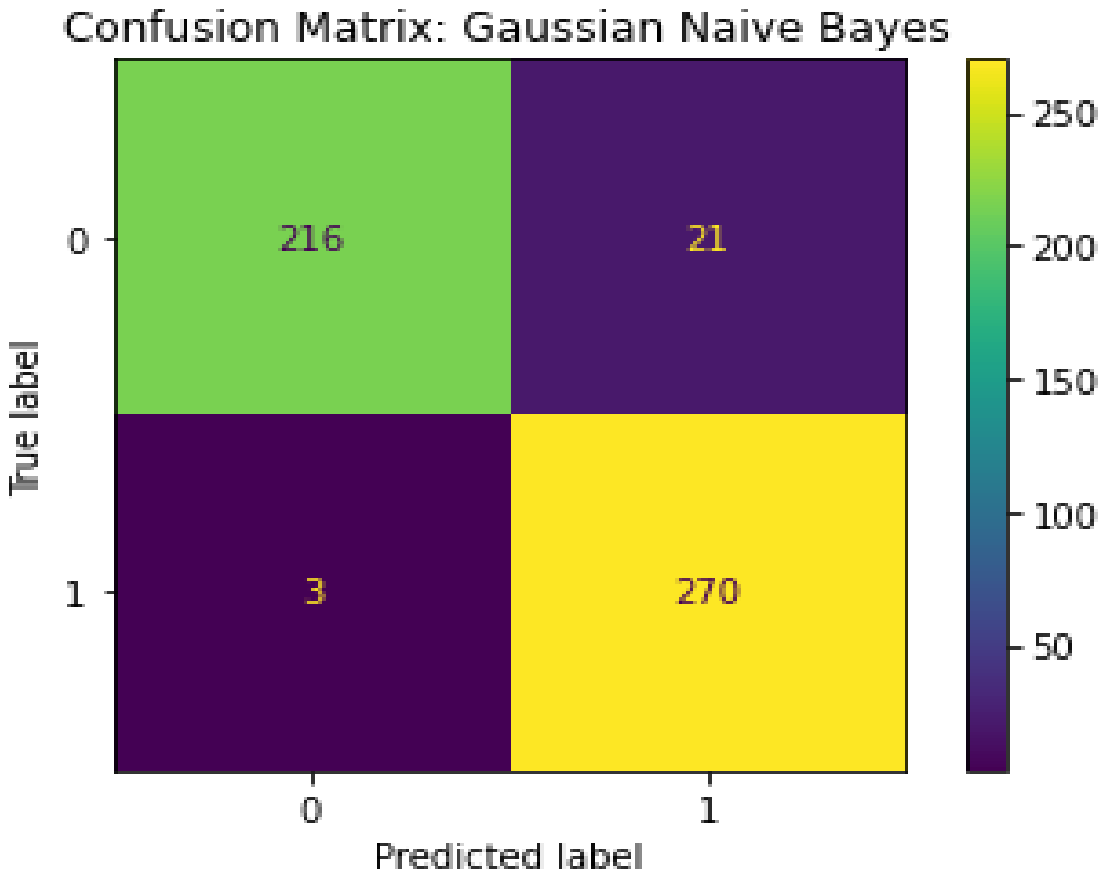


Figure 4.3: Confusion Matrix of Naïve Bayes Classifier

Classification Report of Naive Bayes:

We are able to observe the values for our recall, precision, and f1-score in Table 4.4.

	Precision	Recall	F1-Score
0	0.99	0.91	0.95
1	0.93	0.99	0.96

Table 4.4: Classification Report of Naïve Bayes

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 99 percent.
After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 96 percent.

4.2.4 Logistic Regression

Confusion Matrix:

As can be seen in Fig. 4.4, the True Negative (TN) value is 216, and the False Positive (FP) value is 21. The False Negative (FN) for the second row is set at 3, while the True Positive (TP) is set at 270.

Classification Report of Logistic Regression:

We are able to observe the values for our recall, precision, and f1-score in Table 4.5.

	Precision	Recall	F1-Score
0	0.99	0.91	0.95
1	0.93	0.99	0.96

Table 4.5: Classification Report of Logistic Regression

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 99 percent.
After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 96 percent.

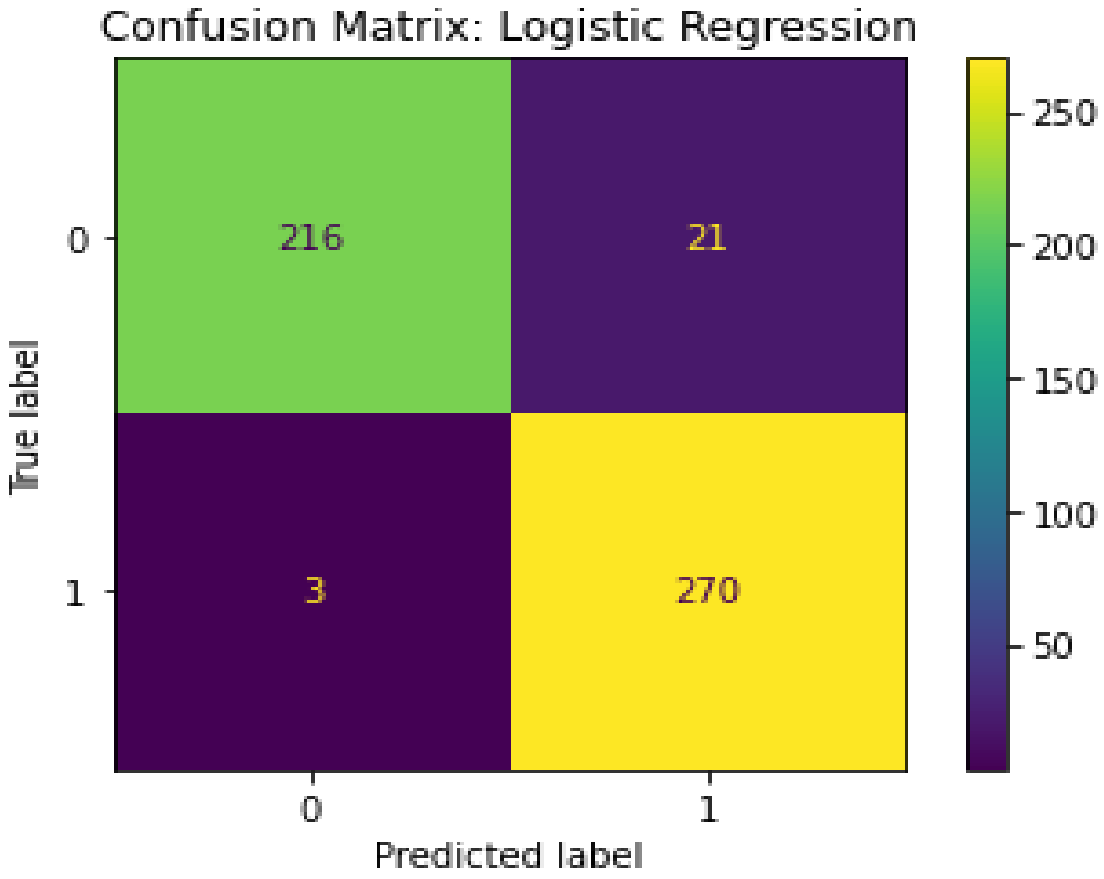


Figure 4.4: Confusion Matrix of Logistic Regression

4.2.5 Support Vector Machine

Confusion Matrix:

As can be seen in Fig. 4.5, the TN value is 216, and the FP value is 21. The FN for the second row is set at 3, while the TP is set at 270.

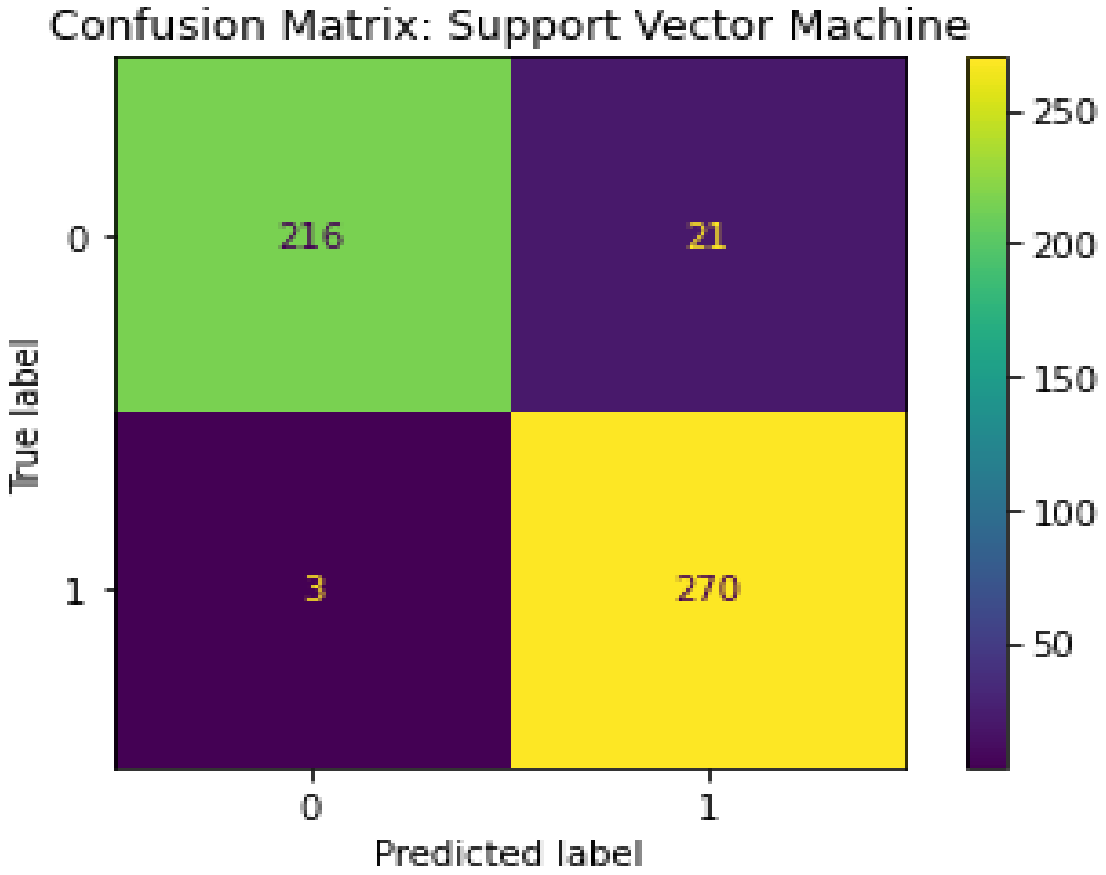


Figure 4.5: Confusion Matrix of Support Vector Machine

Classification Report of Support Vector Machine:

We are able to observe the values for our recall, precision, and f1-score in Table 4.6.

	Precision	Recall	F1-Score
0	0.99	0.91	0.95
1	0.93	0.99	0.96

Table 4.6: Classification Report of Support Vector Machine

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 99 percent.

After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 96 percent.

4.2.6 Gradient Boosting Classifier

Confusion Matrix:

As can be seen in Fig. 4.6, the True Negative (TN) value is 216, and the False Positive (FP) value is 21. The False Negative (FN) for the second row is set at 3, while the True Positive (TP) is set at 270.

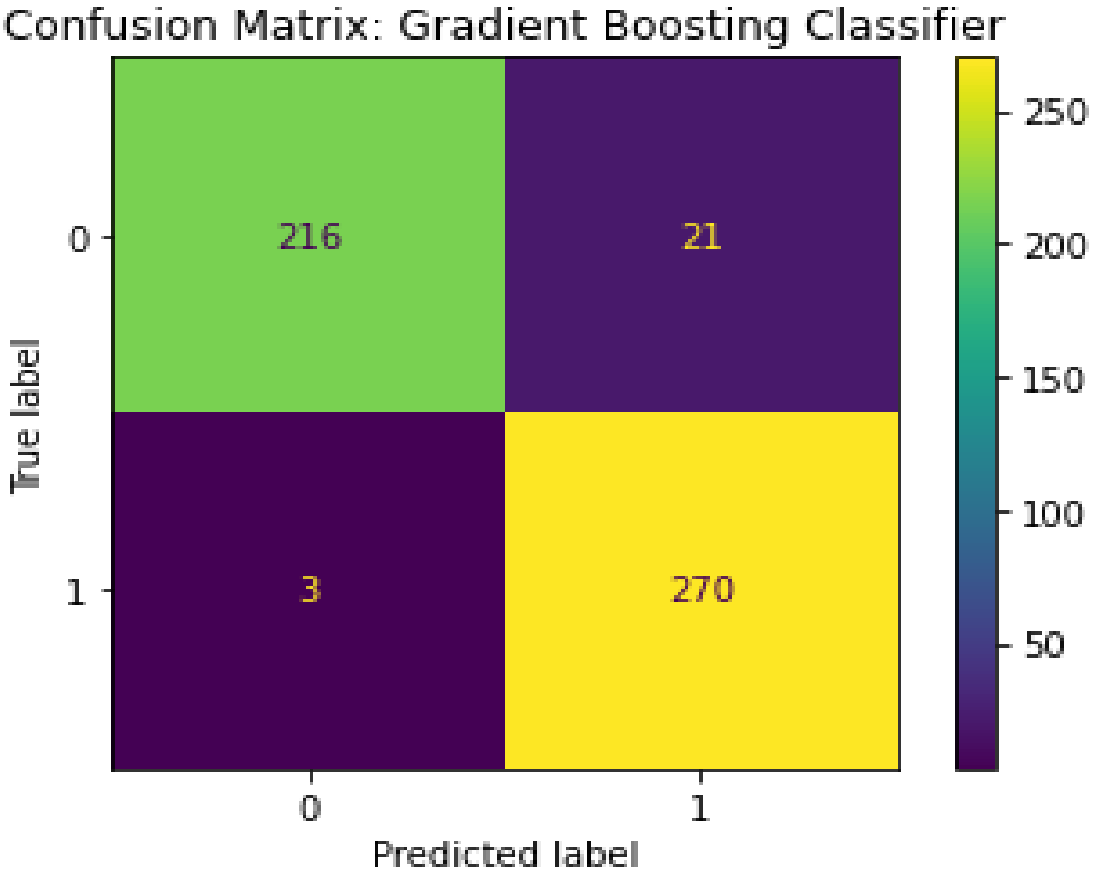


Figure 4.6: Confusion Matrix of Gradient Boosting Classifier

Classification Report of Gradient Boosting Classifier:

We are able to observe the values for our recall, precision, and f1-score in Table 4.7.

	Precision	Recall	F1-Score
0	0.99	0.91	0.95
1	0.93	0.99	0.96

Table 4.7: Classification Report of Gradient Boosting Classifier

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 99 percent.
After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 96 percent.

4.2.7 AdaBoost Classifier

Confusion Matrix:

As can be seen in Fig. 4.7, the True Negative (TN) value is 216, and the False Positive (FP) value is 21. The False Negative (FN) for the second row is set at 5, while the True Positive (TP) is set at 268.

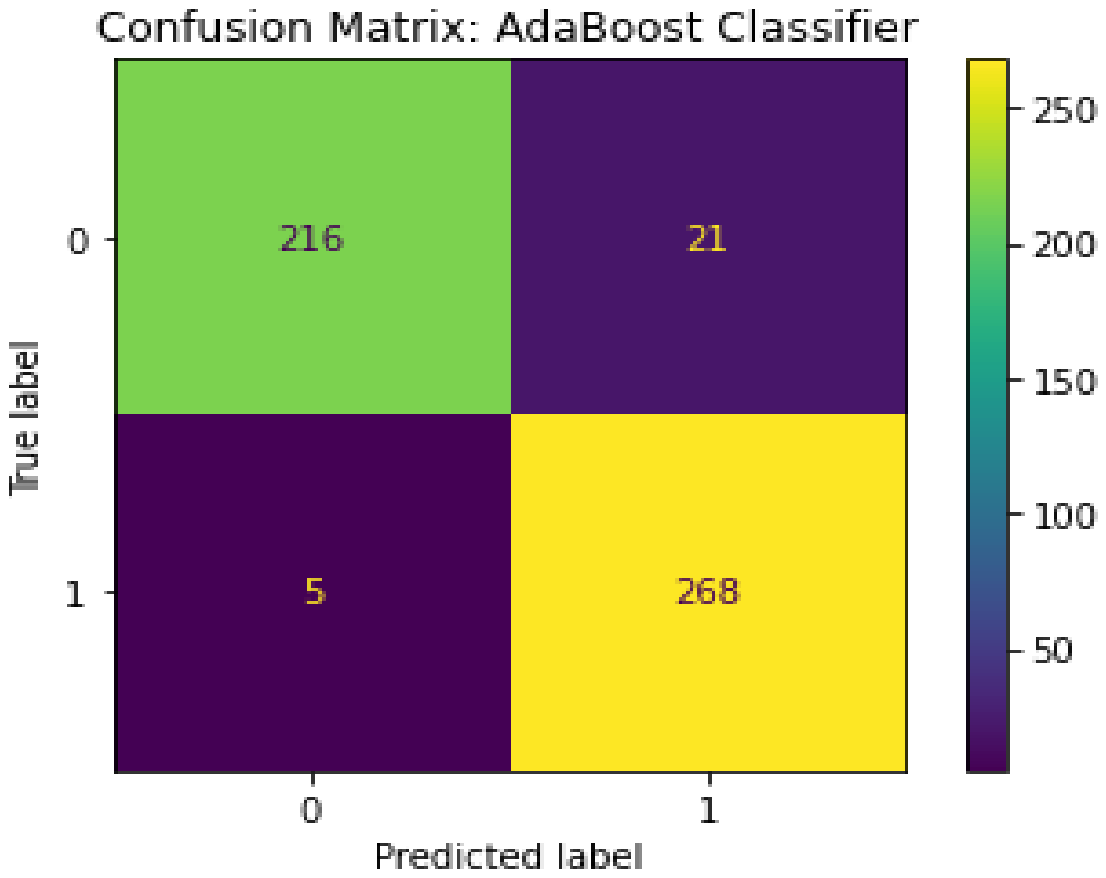


Figure 4.7: Confusion Matrix of AdaBoost Classifier

Classification Report of AdaBoost Classifier:

We are able to observe the values for our recall, precision, and f1-score in Table 4.8.

	Precision	Recall	F1-Score
0	0.98	0.91	0.94
1	0.93	0.98	0.95

Table 4.8: Classification Report of AdaBoost Classifier

Also, if we take a closer look at the precision, wherein precision is defined as the total number of TP in all of the predictions of yes, we can see that the model is quite accurate. We obtain a precision of roughly 93 percent.
Next, we take a closer look at the recall, which is the percentage of actual yes outcomes that correspond to genuine positives. We received a recall rate of roughly 98 percent.
After that, we have a look at the F1-score, which is a measure that takes the harmonic mean of both the recall and precision of a classifier and integrates them into a single value. Its primary purpose is to evaluate the relative effectiveness of two different classifiers. Our F1 score comes in at about 95 percent.

4.3 Comparison Between Different Algorithms

We will therefore examine the precision, accuracy, recall, f-measure of almost all of the methods that were applied to those selected features, including the FN and FP rates for each method. The comparability of all of the algorithms, as shown by a histogram, is presented Fig.4.8 below.

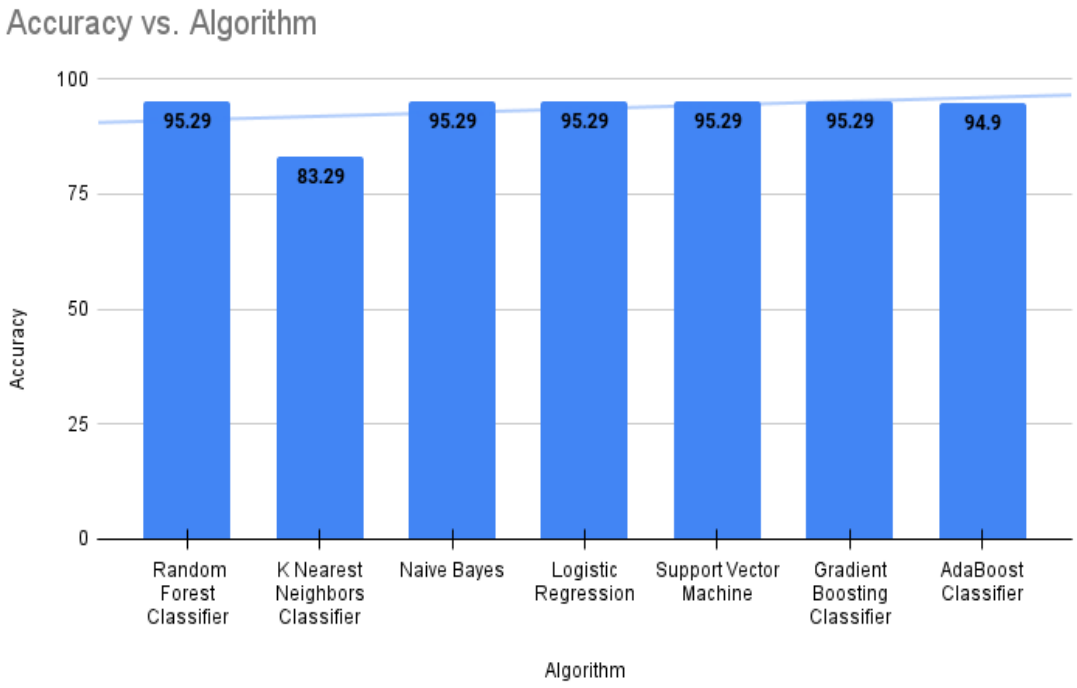


Figure 4.8: Comparability of All Algorithms (Accuracy)

From the above graph, we can easily find that 95% is the highest accuracy. Therefore, it is clear to see that the findings that we obtained after analyzing our dataset using

a variety of methods are, for the most part, comparable. Random forest classifier, Naïve Bayes, Gradient boosting classifier, Logistic Regression and Support Vector Machine come out on top as providing the perfect effectiveness of all the methods.

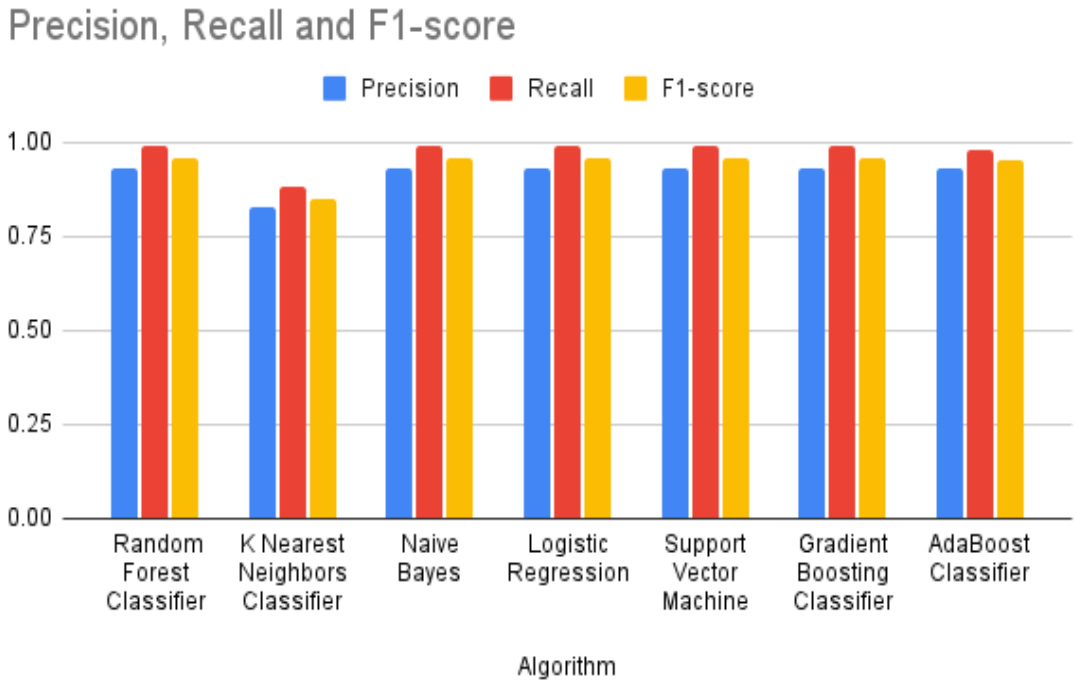


Figure 4.9: Comparability of All Algorithms (Precision, Recall, F1-score)

Furthermore, we can also observe that False Positive and False Negative have a nice equilibrium. We can however conclude from the Fig. 4.9 precision, recall, and f-measure that Random Forest Classifier, Naive Bayes, Logistic Regression, Support Vector Machine, and Gradient Boosting Classifier provide the best possible results for our model in both scenarios. To summarize, we may reach the conclusion that employing the recommended 16 features is viable for our model based on the comparison of accuracy, precision, recall, and the f-measure value. This not only prevents the data from being overfit, but it also makes the dimensions of the data more manageable. As we can see from the performance compared of their accuracy, they displayed an even higher level of performance. Since we can see that Gradient Boosting Classifier, Support Vector Machine, Random Forest Classifier, Naive Bayes, Logistic Regression have the highest f-measure, we will be using these five algorithms in the coming future and will continue to explore and eventually expand our work so that we can improve our accuracy and assist policymakers in making the best decisions for the fast-growing hotel industry.

Chapter 5

Final Remarks

5.1 Conclusion

Hotel industry is among the most potential industries for industrial prosperity in underdeveloped countries all over the world. Bangladesh isn't the exception rather than the general rule. Bangladesh is trailing behind in the hotel business when compared to its neighboring countries.
Using an online questionnaire and an analysis of the data, we set out to discover the quickest and most efficient Machine Learning algorithm for forecasting the purchase of a hotel package. This research analyzes the accuracy scores of the Random Forest Classifier, K Neighbors Classifier, Naive Bayes, AdaBoost, Support Vector Machine, Logistic Regression, and Gradient Boosting algorithm for the prediction of the purchase of a hotel package. This study discovered that the Gradient Boosting Classifier, Support Vector Machine, Random Forest Classifier, Naive Bayes, Logistic Regression with an accuracy score of 95 percent, are by far the most effective algorithms for forecasting the purchase of a hotel package, according to the findings. If we were to construct a software based on our model that included additional data from a larger dataset than the one, we used, we may make significant improvements to our work in the future. Moreover, it will aid in the identification of current trends in this industry throughout the world, while also portraying the existing situation in Bangladesh including its future potential. Future initiatives to be taken, as well as proposals for long-term economic development in a sustainable way, are discussed in detail.

5.2 Limitations and Future Work

We tried our hardest to make our work a higher accuracy system, but it still has certain shortcomings, as you can see below. Limitations:

- A total of 2550 observations were found in the dataset we used for this analysis. As a result, our size of the sample, training set, and test dataset are all on the conservative side. The greater the size of the dataset, the more accurate the prediction will be. If we utilize a large - scale dataset than the one we've used so far, the accuracy of our suggested technique will be significantly improved, and our model will become much more standardized as a result.

- The number of selected factors is so limited that we have been unable to include less significant relevant attributes in our survey questionnaire because doing so would lengthen our research paper and jeopardize the collection of data process' legitimacy.
- Another limitation was that, as a result of the current pandemic situation, we have been unable to reach people in order to conduct this questionnaire.

In the future, we hope to improve the accuracy of our model and identify more important reasons or characteristics that are associated with the hotel industry to help policy makers to make better decisions. At this time, we are attempting to collect further information from online sources. We will include more types of data in our research in order to broaden our scope. We will strive to include more approaches that concentrate a larger emphasis on data preprocessing into our workflows in the future. We'll try to obtain more information by doing an offline survey, which is now impossible due to the outbreak, in order to obtain more information. Machine learning techniques will be deployed in greater numbers in order to improve our performance. We focused all of our attention on a single country, which was Bangladesh. In the future, we hope to work on databases that include data from other nations.

Bibliography

[1] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580–585, 1985.

[2] R. E. Wright, "Logistic regression.," 1995.

[3] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[4] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve bayes.," *Encyclopedia of machine learning*, vol. 15, pp. 713–714, 2010.

[5] O. Kramer, "K-nearest neighbors," in *Dimensionality reduction with unsupervised nearest neighbors*, Springer, 2013, pp. 13–23.

[6] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in neurorobotics*, vol. 7, p. 21, 2013.

[7] R. E. Schapire, "Explaining adaboost," in *Empirical inference*, Springer, 2013, pp. 37–52.

[8] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[9] P. Lison, "An introduction to machine learning," *Language Technology Group (LTG)*, vol. 1, no. 35, pp. 1–35, 2015.

[10] A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Information Processing in Agriculture*, vol. 3, no. 4, pp. 215–222, 2016.

[11] M. Kuhkan, "A method to improve the accuracy of k-nearest neighbor algorithm," *International Journal of Computer Engineering and Information Technology*, vol. 8, no. 6, p. 90, 2016.

[12] N. Antonio, A. de Almeida, and L. Nunes, "Predicting hotel bookings cancellation with a machine learning classification model," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2017, pp. 1049–1054.

[13] A. Bronshtein, "A quick introduction to k-nearest neighbors algorithm," *Noteworthy-The Journal Blog*, 2017.

[14] M. Nilashi, K. Bagherifard, M. Rahmani, and V. Rafe, "A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques," *Computers & industrial engineering*, vol. 109, pp. 357–368, 2017.

[15] T. Srivastava, "Introduction to k-nearest neighbors: Simplified (with implementation in python)," *analytics vidhya*, vol. 26, no. 03, 2018.

[16] A. Ahani, M. Nilashi, O. Ibrahim, L. Sanzogni, and S. Weaven, "Market segmentation and travel choice prediction in spa hotels through tripadvisor's online reviews," *International Journal of Hospitality Management*, vol. 80, pp. 52–77, 2019.

[17] A. Jeyaratnam, I. Mahakalanda, and T. De Silva, "Travel demand analytics based customer-service decision model," 2020.

[18] A. Martínez, C. Schmuck, S. Pereverzyev Jr, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, 2020.

[19] A. Bansal and P. Srivastava, "Factors affecting consumer buying behavior of online travel agencies," *Elementary Education Online*, vol. 20, no. 1, pp. 2958–2958, 2021.

[20] L. N. Pereira and V. Cerqueira, "Forecasting hotel demand for revenue management using machine learning regression methods," *Current Issues in Tourism*, pp. 1–18, 2021.

[21] S. Oh, H. Ji, J. Kim, E. Park, and A. P. del Pobil, "Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service," *Information Technology & Tourism*, vol. 24, no. 1, pp. 109–126, 2022.

[22] M. Sultana and M. S. Islam, "Global tourism trends and bangladesh tourism: A nexus,"