Exploring the Influence of Social Media Engagement on Online Shopping Preferences of Youth and Young Adults: A Machine Learning Approach

by

Md Faysal
19101504
Abu Zawad Ahmed
19301061
Abdul Karim Sohag
23141061
Razia Rahman Maria
24341113

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
October 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md Faysal 19101504	Abu Zawad Ahmed 19301061
Abdul Karim Sohag 23141061	Razia Rahman Maria 24341113

Approval

The thesis/project titled "Exploring the Influence of Social Media Engagement on Online Shopping Preferences of Youth and Young Adults: A Machine Learning Approach" submitted by

1. Md Faysal(19101504)
2. Abu Zawad Ahmed(19301061)
3. Abdul Karim Sohag(23141061)
4. Razia Rahman Maria(24341113)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October 20, 2024.

Examining Committee & Signature:

Supervisor:
(Member)

Md. Sabbir Ahmed
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The modern-day rise of social media platforms and their influence on customer behavior has been a trending topic. It is important for businesses to understand the connection between young people's engagement with social media and their online shopping habits. This research paper aims to provide how engagement with social media affects an individual's online purchase decision. This study helps us learn how social media and influencers have an impact on today's youth. For this purpose, we derived 4 sections of online interactions and shopping i.e., General Engagement, Online Food Purchases, Online Subscription Services, and Online Gaming. The objective of this research is to understand the online shopping preferences of young people by statistical analysis of their social media activity, such as likes, engagement, sharing, and interactions with influencers and followers. Data has been gathered through conducting a curated survey among the youth and young adults. A total of 1018 data has been collected for the research through online survey forms which were utilized for running machine learning models. The outcomes of the study shed insight into how young people's online buying habits are influenced by their use of social media. It also concluded how high social media engagement ensues high online purchase behavior. Businesses may improve their marketing tactics by creating more social media-based ad campaigns which will be able to reach more modern-day youths. Influencers and creators can leverage this study to increase their reach and capitalize on influencer marketing strategies. Additionally, this research provides consumers with market information, which will in turn encourage a degree of caution when faced with social media ads and marketing strategies designed by businesses.

Keywords: Online Shopping; Social Media; Prediction; Machine learning; Statistical Analysis; Consumer Pattern and Behaviour;

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

In the modern world more or less every individual is familiar with the Internet and uses it for various purposes. One of the notable ones that people use it for is to make online purchases. The action to buy or sell products through an online method is commonly known as online shopping. The way individuals purchase and behave as consumers has undergone amazing changes in the current era compared to a few years back. The younger generation is among those leading this digital revolution, especially teens and young people who are heavily immersed in social media and online activities. Thus, they are the prime candidates for participating in activities such as online shopping. This thesis sets out on a journey to investigate an interesting phenomenon within this constant digital favoring time: how young customers' preferences and social media participation affect their online buying behaviors. This intersection of two fields is a fascinating topic for research, shedding light on the shifting dynamics of consumer behavior in our present age. Several important considerations highlight how important this research is. First off, as a population, young people have a big influence on the consumer market. Zheng Xu et al.[10] in their article talked about 4 important characteristics, satisfaction of customers and the merits and demerits of strategies for the marketing field through quantitative analysis and to put forward reasonable marketing tactics, opinions on improvement and suggestions to raise customer satisfaction.

In addition to influencing current market trends, their decisions and actions also provide hints about how business will change in the future. Because of this, businesses and marketers need to comprehend their online purchasing behaviors, especially in light of how they use social media.

Furthermore, the prevalence of social media platforms in young people's lives offers a rare opportunity to research how these platforms affect their purchasing behavior. Nowadays, social media is becoming more than just a tool for communicating. It also serves as a portal for discovery, an inspiration source, and an encouragement for user involvement. Investigating how their social media behavior affects their online purchase decisions, therefore, offers to yield priceless information for the current world's economy. The importance of machine learning as a useful tool is emphasized throughout this research. The complex nature of this study may be uncovered using machine learning, which is renowned for its ability to analyze big data sets and identify subtle patterns. This study uses machine learning algorithms to mine the massive amounts of data produced by social media interactions and online purchasing activities for useful information.

1.2 Problem Statement

In the age of science social media and its effects are real. Now people are more relying on online platforms for their needs. Young adults are the biggest community who prefers the online platforms most. Here we are trying to work with the biggest community specifically so that they get more relevant products for their wants and their likes and find what and how social media is influencing their choice in terms of shopping. Our goal is to find the impact of modern-age social media influence on the youth generation in a particular age group and how it's creating differences and impacting their decision-making. We will move to the machine learning model methods and algorithms to determine the outcome. We will be collecting data on young adults to find and determine the thought processes and come to a conclusion. Different types of solutions and paths will be explored in order to find a more accurate and effective solution. Combining with a machine learning approach a better algorithm will be implemented to suggest more accurate products according to the tests and needs of young people which will contribute to filtering out irrelevant products more.

1.3 Objectives

1) Analyze how social media affects consumer behavior: Examine how consumer behavior has expanded as a result of social media's involvement in daily life, paying close attention to the requirements of youths and young adults in particular. Recognize how consumer contacts, engagements, and preferences are impacted by social media involvement.

2) Use Theoretical and Experimental Methods to Gain Understanding: Gain insight into how marketing practices affect client views by combining experimental and theoretical methodologies. To better understand client satisfaction and to offer suggestions for improving marketing strategy, combine qualitative and quantitative analyses

3) Analyze the way Influencers Promote Trends and gain Consumer Attention: Social media has a lot of contributions on promoting brands. The social media content creators known as influencers deliver content related to brand promotion and spread new trends among their followers. This way the strategies of celebrity content creation make the brand more appealing and approachable to the consumer.

4) Use Machine Learning to Predict Consumer Behavior: Create predictive models utilizing machine learning techniques to analyze the enormous amounts of data generated by social media platforms to gain valuable knowledge about the patterns and trends in consumer behavior. Analyze how well machine learning can be used to predict and understand what consumers want.

Chapter 2

Related Work

In the thesis paper by Ebrahimi et al.[8] it can be seen that using the convenience sampling method, relying on five factors (e.g. entertainment, customization, interaction, word of mouth, and trend), the collected data demonstrated a positive influence of all five factors on people's buying behavior. Out of the 475 surveys distributed online, 466 people completed them, resulting in a response rate of 98.1%. The study revealed that factors such as entertainment, customization, interaction, word of mouth, and influencer marketing on social media have a positive impact on buying behavior. This insight suggests that marketers can utilize these factors to shape customers' purchasing decisions.
According to the published research paper Hossain & Sultana[15], in online business, it's essential to understand why customers choose to buy products online and why they choose not to. Customer behavior can vary based on factors like age, gender, and more. According to the findings, exceptional discounts/offers stimulate purchases in Bangladesh, while a lack of trust inhibits them. After analyzing the data, it was found that random forest achieved a high accuracy of 97.89% compared to other machine learning approaches.
Researched and published by Jothi & Gaffoorcite[1] social media platforms influence consumer behavior in virtual environments, especially in terms of message exposure and the connection between the provided information and a potential buyer's decision to make a purchase. Furthermore, it's evident that individuals in the age group 18-22 are more inclined to purchase products through online shopping. Thus, researchers and marketers alike are positioned to gain valuable insights into the dynamics of consumer behavior throughout social media networks by leveraging the capabilities of machine learning.
In a research paper by Chaudhary, K., Alam, M., Al-Rakhami, M. S., & Gumaei, A. (2021)[5] published in the Journal of Big Data, researchers delve into the fascinating intersection of advanced technologies like big data analytics and machine learning to predict how consumers behave on various social media platforms. This study primarily focuses on platforms such as Facebook, LinkedIn, Twitter, YouTube, Instagram, and Pinterest. The main goal of this research is to create a predictive model that can anticipate and understand how consumers interact with content, products, and services on these social media platforms. To achieve this, the researchers followed a step-by-step process that involved cleaning up and transforming the raw data. This pre-processing stage is crucial to ensure the data is accurate and can be used effectively. A crucial part of the study involves comparing the performance of different

machine learning models. The researchers used two key measurements to assess the models: root mean square error (a measure of prediction accuracy) and overall accuracy. These metrics help determine which model can best predict consumer behavior accurately. Interestingly, the Decision Tree Regression model emerged as the most effective, showing the least prediction errors and the highest accuracy rates. This model excelled in identifying complex patterns within consumer data. By showcasing the effectiveness of the Decision Tree Regression model, the research highlights the power of machine learning in understanding and forecasting consumer behavior within the ever-changing landscape of social media. The findings emphasize how data analytics can extract valuable insights from large data sets, enabling businesses to optimize marketing strategies, improve user engagement, and enhance the overall consumer experience.

The article written by Zheng Xu et al.[10] uses a combination of experimental and theoretical methods to find out which marketing behavior of businesses affect customers' opinion of choosing that business' product. This research has five components of marketing behavior factors, those are namely non-standard promotion, integrity, social responsibility, competition and technical behavior. The study survey tackled 4 main aspects. Those being customer approval, an overview of customer satisfaction and the merits and demerits of marketing tactics through quantitative analysis and to put forward reasonable marketing strategy improvement opinions and suggestions to improve customer satisfaction. They used the ELLA algorithm.

The paper authored by Satpathy el al[2] describes a machine learning technique with a Neuro-Fuzzy system focusing on particular fields like business, education, society, and youth and how it broadly affects society. As the number of media users is rising with the online pouches sites, it becomes necessary for the major brands to understand the customers' behavior. In this work, they tried to reveal the hidden parameters that affect the purchasing behavior of a customer. As these are characterized by uncertainty, they will use Fuzzy Logic broadly.

Arasu et al.[3] mentioned in their paper that they used a machine learning approach with extracted data by data mining technique. They used the WEKA(Waikato Environment for Knowledge Analysis) algorithm. WEKA is mainly a machine learning tool that uses different algorithms for different scenarios. This can achieve better performance than other machine learning tools.

In the published thesis paper composed by Madani & Alshraideh [7] examined the use of machine learning models for predicting consumer purchase decisions in Bangalore, India. A dataset of 55 variables collected from 388 customers is used, including customer preferences and demographic information. The research showed the usage of four prediction models. These are classification and regression trees (CART), C4.5 decision trees, random forest, and a rule-based classifier. Among the performance of predicting purchasing decisions of the four models, the C4.5 decision tree performed the best with an accuracy of 91.67%. The dataset focused on predicting buying decisions by analyzing a variety of characteristics, demographic details, and customer preferences. The study revealed how ease and convenience, good taste, time-saving, and influence of restaurant ratings affect customer's decision highlighting the importance of predictive models and the application of machine learning for predicting consumer behavior in the online food delivery service.

The research paper written by Parihar & Yadav [9] examines predicting the behavior of online consumers to improve marketing strategies in online businesses. The methodology of the research uses both numerical and categorical elements to analyze a dataset that has 18 features that were collected from around 12330 sessions of different users over one year. However, The methodology includes exploratory analysis, correlation assessments, and detailed examinations of web pages, page metrics, visitor information, and visit dates. Based on performance metrics various classification algorithms, including Naive Bayes, Logistic Regression, Random Forest, and Gradient Boost are implemented and evaluated in it. The report also reviews revenue trends related to visit dates, special days, weekends, and weekdays. Again, the comparative analysis of classification models highlights Gradient Boost as the best-performing algorithm as it has a precision of roughly 76% and an accuracy of about 91% for predicting whether a customer will make a purchase or not, followed by Random Forest, SVM, and others. The paper focuses on the importance of accuracy in predicting customer purchase behavior for effective business strategies.

From the research papers that we have reviewed, we have learned about important factors that have an impact on online purchase decisions. We came to learn how social media engagement, entertainment, interaction, discounts, and influencers can affect the purchasing decisions of users. We have seen that most of the users of social media platforms are young users. And they are greatly influenced by social media to make online purchases. Also to predict purchasing behavior different machine learning models like Random Forests, Decision Trees, and Regression models have been used. Thus, these studies provided us with approaches that show how social media impacts digital marketing and purchasing decisions.

Chapter 3

Methodology

First, we collected data through surveys based on three different segments including online food purchases, online subscriptions, and online game purchases. Then we cleaned our datasets for quality data and went for data preprocessing. We also did statistical analysis to prove our hypotheses which were based on the online buying behavior of consumers that was impacted by social media engagement. The gathered datasets that were applied to our code were then split for train and test sets for each segment. After that, we selected and implemented machine learning models to predict consumer behavior.

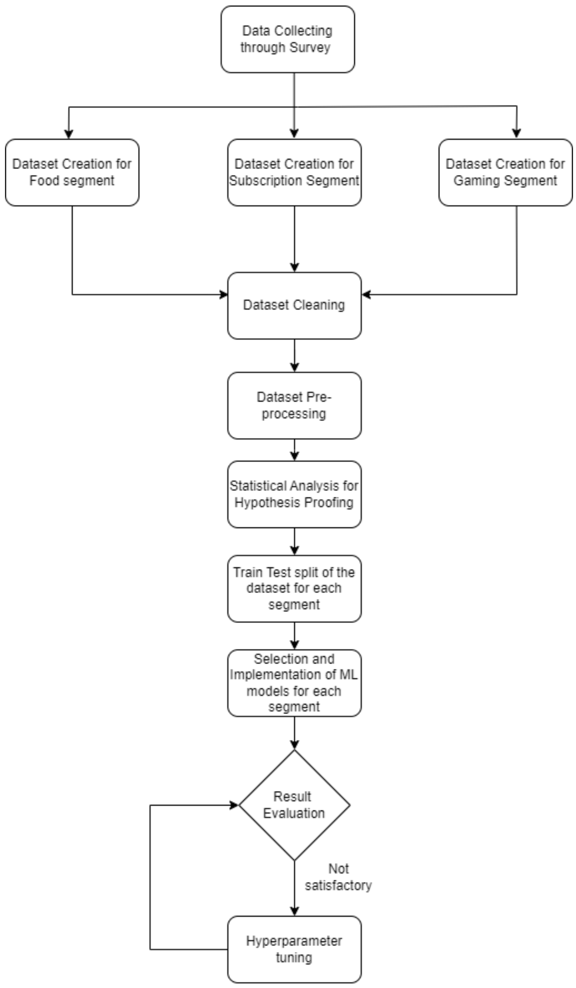

Figure 3.1: System Overview

3.1 System Overview

We start with collecting data. As this type of data wasn't readily available, we had to face a few issues due to the lack of datasets. That's why we had to make our own dataset by surveying people. We started considering three specialized fields in the online sphere, those being: online food purchase behavior, online subscription behavior, and online game purchase activity. We also included two other sections: general information and social media engagement, which were given to gauge the impact and engagement of our targeted group of people. We made three different datasets for each of these three specialized fields. Then, we cleaned them by deleting unnecessary columns of data and started to pre-process the data. After pre-processing we did statistical analysis for hypothesis proofing. Then, we did a train test split of the dataset for each segment. After that, we implemented machine learning models for each segment and evaluated the results. If the result does not satisfy, we will use hyperparameter tuning and evaluate the result again. Thus, it allowed us to initialize and count it as a satisfactory result. Our central database was updated with the recent progress and outputs. And thus we were able to get the desirable and satisfactory outcome with the process coming to an end.

3.2 Model Description

We particularly chose specific models like Random Forest, Decision Trees, Gradient Boosting, XGBoost, Naive Bayes and Ensemble Model so that we can achieve accurate results for our predictions on online purchasing behaviors. Each model handles the dataset differently and these models are the best suited for our dataset. So we have applied multiple models to understand which one provides the best outcome depending on data complexity and model interpretability and their performance.

3.2.1 Naive Bayes Classifier

To do the classification tasks in supervised learning, Naive Bayes Algorithm is very effective. Naive Bayes is based on the Bayes theorem. In text classification, this model performs well. It is faster than other classifiers. It is also easy to implement and suitable for large datasets. It can also handle a large number of features. As in our dataset, there are many features, and this classifier is suitable for it. However, the Naive Bayes Classifier calculates probabilities independently for each feature. As a result, if any feature has some missing data, it does not cause problems in the calculation of other features. Again, it requires less training data than others and the response time of this classifier is less than other classifiers. To work with classification tasks Naive Bayes works very well. It predicts or classifies using probability theory. Firstly, the Naive Bayes classifier will be implemented in the features of our dataset. Then the overfitting will be checked by using the K-fold cross-validation process. Where the dataset will be divided into K subsets and will be trained K times.

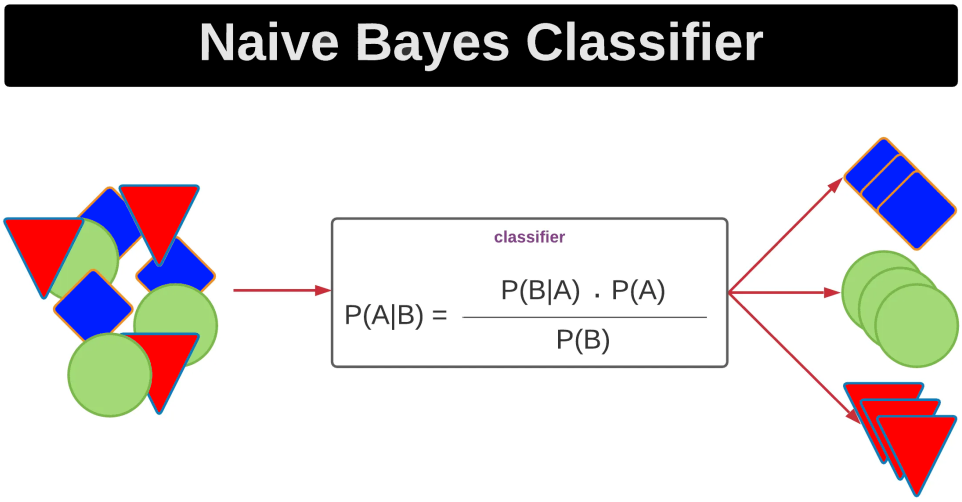

Figure 3.2: Naive Bayes Classifier
[13]

3.2.2 Gradient Boost

A gradient boost-based algorithm is XGboost Algorithm, which is very fast compared to other algorithms. Because It uses the parallelization concept and can optimize the cache very well. Again, it can optimize the memory in such a way that it can work on data that is larger than the size of the RAM. Moreover, it can regularize tasks, which prevents the model from overfitting. In the normal gradient boost algorithm, there is no regularization. This algorithm is very popular with data scientists and this algorithm has lessened many challenges of Machine learning. Gradient boost classification is good at multi-task classification. Firstly, we will construct a base model. In base model calculation, a probability will be calculated with the values of dependent variables. Then we will count the residual errors. And construct a decision tree considering the input and output. Then the data will be trained by that decision tree.

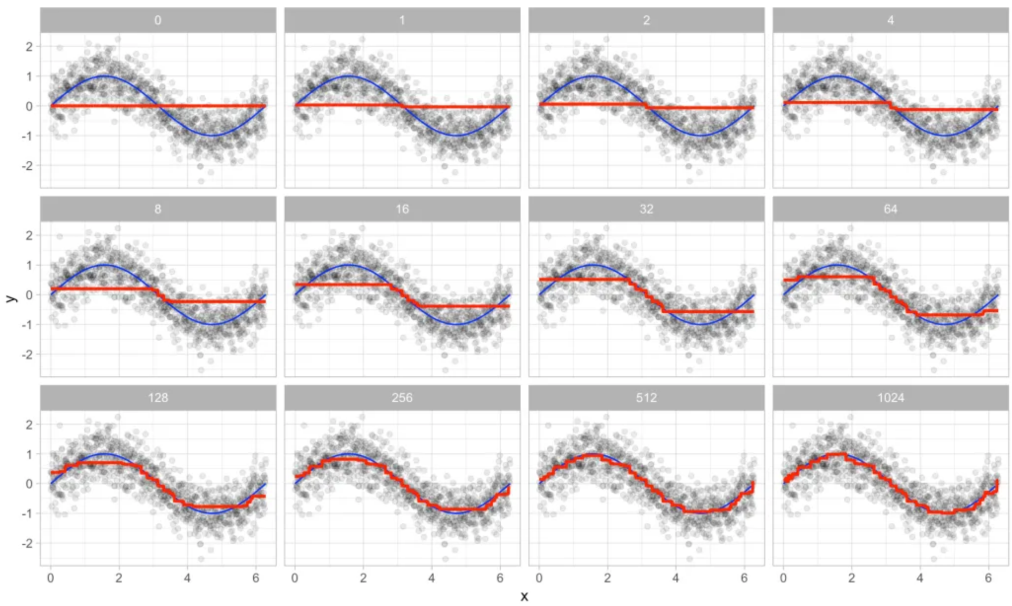

Figure 3.3: Gradient Boost
[14]

3.2.3 Random Forest

Random forest is a widely known machine learning algorithm. Problems of classification and regression is solved by this mechine learning method. It is an ensemble of decision trees that works by adding multiple decision trees in the time of training. The result is measured by the majority of voting for classification problems and takes the average value for regression problems. The model captures various samples by combining multiple trees and thus it gives more accuracy. It works well for nonlinear relations in the data. This algorithm reduces overfitting as it selects features randomly. Random forest improves bagging in random subsets by de-correlating trees with the introduction of a split of features in random subsets which makes it easier to work with only small subset features but not the whole features of the model. Random forest has some advantages. It can reduce the possibility of overfitting and also lowers the amount of training time. It can work well with big data sets and can create accurate predictions. We can get a high level of precision from this model and there is also no need or requirement of variables normalization. To implement this model for our research first, we are going to define the target variable and select the features. Then after pre-processing our data, we will be working on the train test segment by dividing the dataset into training to train the model and testing sets to evaluate the performance.

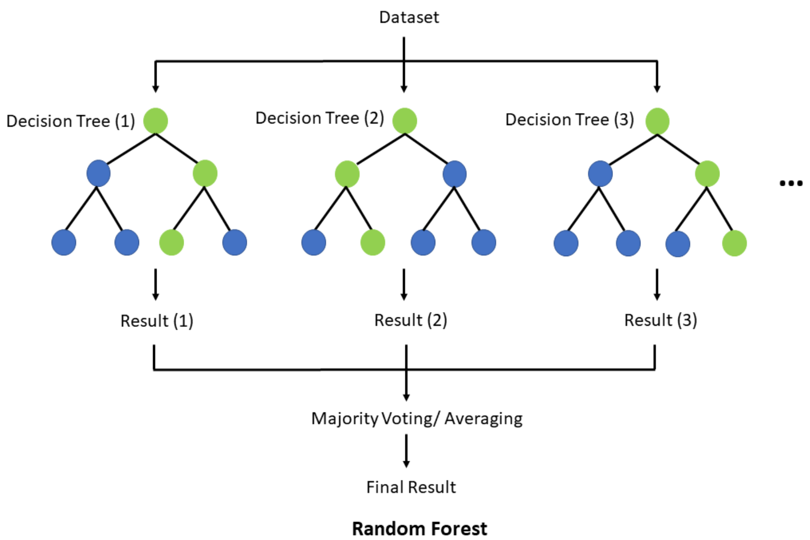

Figure 3.4: Random Forest Classifier
[11]

3.2.4 Decision Tree

Decision Tree is an algorithm that is very easy to understand and it is quite interpretable. The algorithm can work with numeric and categorical data and also it works well with nonlinear relationships in the data. The model has a tree-like structure where the root is based on entropy gain there is an internal node that is the features, the branches are the decision of the features and there is a leaf node that provides the outcome of the target variable. The tree structure helps in making decisions through the number of questions until the result comes. It is a commonly used method of nonparametric administered learning. The method begins with a set of questions and after the questions are answered, it leads to the next question and in this way the tree finalizes the value at the end. The main disadvantage of the decision tree is that it is prone to overfitting and pruning techniques can help to overcome the issue. Decision trees provide clear decisions and continuous segmenting of data based on a particular parameter so to implement this algorithm on our project first we need to select the best features from our dataset then we are going to divide our dataset into subsets for chosen features and after that, we are going to check with K-Fold cross-validation to see if there have been any overfitting happened.

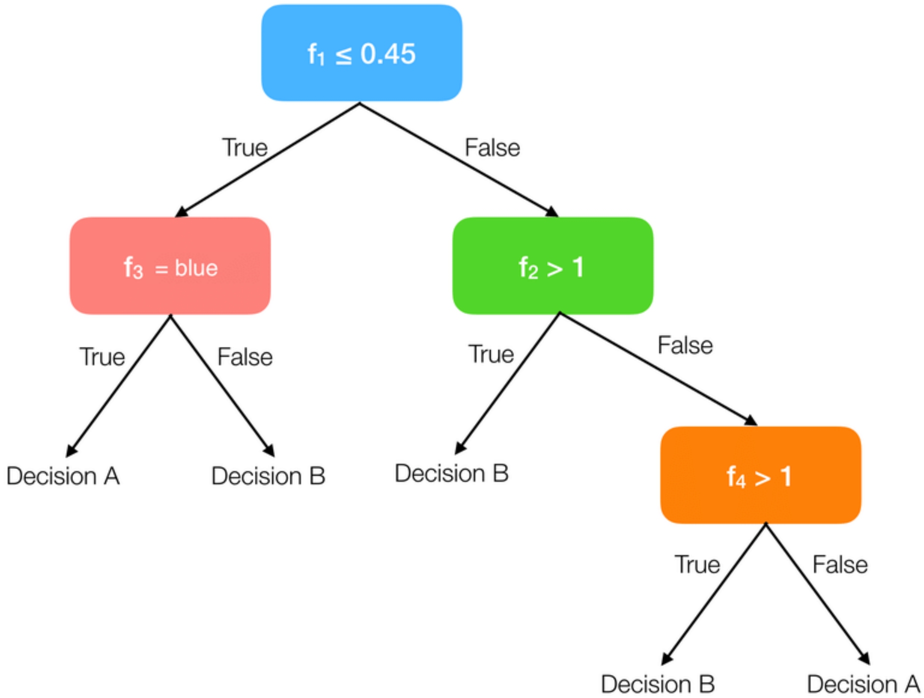

Figure 3.5: Decision Tree Classifier
[6]

3.2.5 XGBoost Classifier

XGBoost is the improved version of the gradient Boost algorithm and is referred to as Extreme Gradient Boosting. Because of the efficiency and better accuracy of the model it has been used in this project. The implementation of the model is highly efficient for speed optimization and performance. This model works faster than other gradient-boosting libraries during parallel processing and it also offers regularization that helps to avoid overfitting issues. The model can cope with missing values while training the dataset. XGBoost maintains an iterative process that can minimize the loss of function because it works by building sequential decision trees where each newly formed tree tries to fix the mistakes of the previous ones. The model first starts by learning a simple model then it calculates the difference between the actual and predicted values so that it can compute the errors, after that to predict the mistakes it fits a new tree and lastly updates the model by adding trees for better accuracy. Until the residual minimization is done the process of computation and updating the model repeats iteratively. The total balanced final prediction is the outcome of all summed individual tree predictions that have been adjusted by the learning rate. XGBoost is a quick model that provides higher accuracy and can manage huge datasets by producing robust predictions with complex relationships among features.

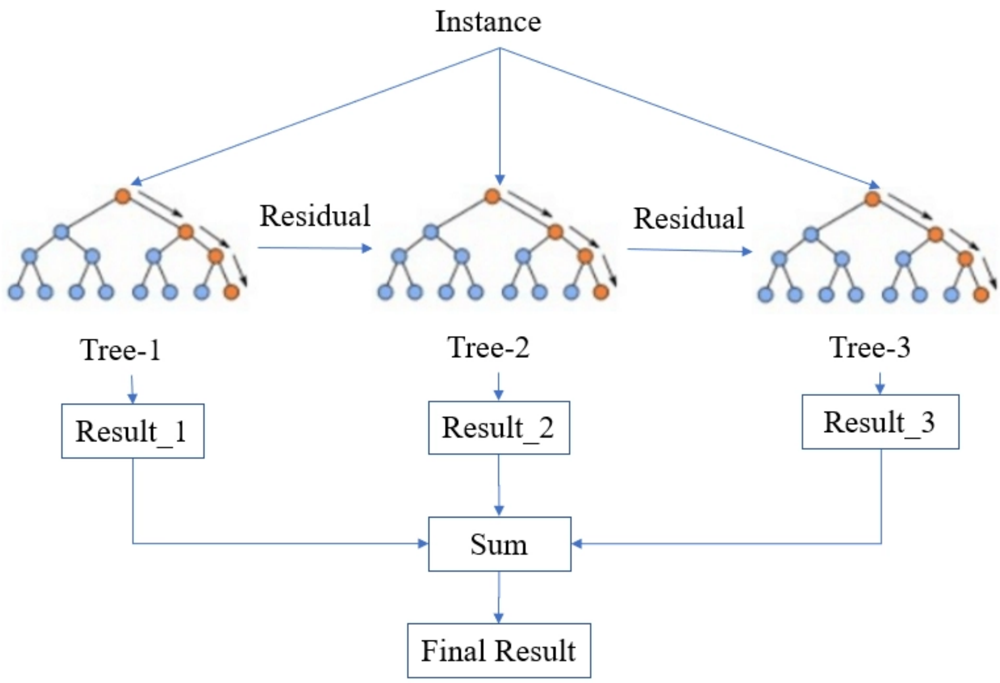

Figure 3.6: XGBoost Classifier
[4]

3.2.6 Ensemble Model Classifier

The ensemble model is used to create a more accurate and robust prediction by combining predictions from other base models that are known as weak learners. The objective of the model is to reduce variance, and bias and improve accuracy in prediction by computing the combined strengths of multiple models. Ensemble Model works in two categories including Bagging and Boosting. In the case of Random forest, it falls under the method Bagging, which is also known as bootstrap aggregating works by reducing variance while training different models separately on various sections of the datasets followed by an average of the predictions of those models. The other method which is Boosting minimizes bias by training the models in a sequential way, with each new model that fixes the mistakes of the older one. XGBoost, Gradient Boosting, and AdaBoost are some ensemble models that fit into the boosting category which, to create strong prediction, uses an ensemble of decision trees. The ensemble model works by enhancing the predictive accuracy and robustness of combined algorithms.

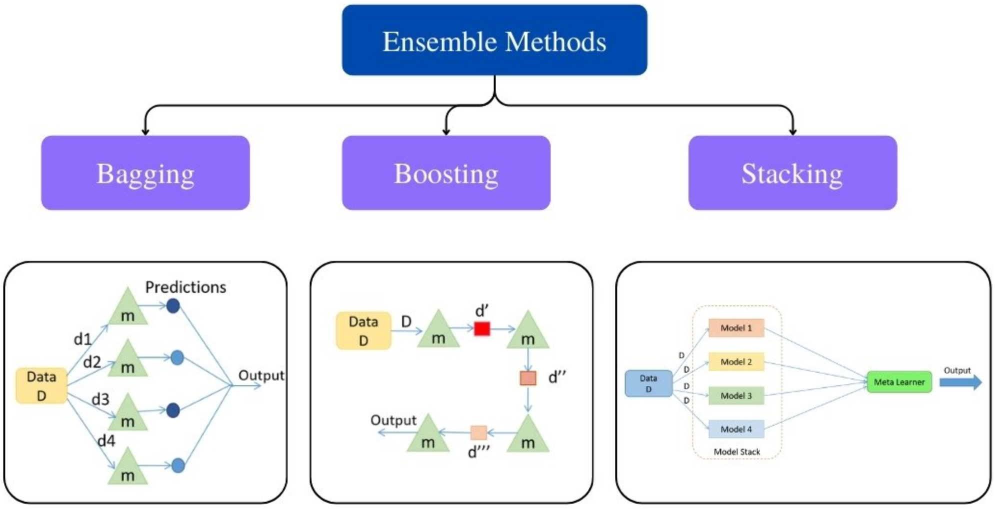

Figure 3.7: Ensemble Model Classifier
[12]

3.3 Data Collection and Preparation

For our research purpose, a 44-question survey was prepared based on our target variables. These include online food purchases, online subscription service, and online gaming/game purchases along with a few general information-related questions. We collected the data from 1018 respondents who answered all those questions. From the questionnaires, we have calculated the percentages of users' involvement in social media activity and their purchasing behavior. Then to maintain the quality of the data we cleaned the datasets and then pre-processed the data. So, that it can be applied to the code for predicting online buying behavior through machine learning models.

3.3.1 Dataset Collection

For selecting features and making a dataset, we followed some criteria or targeted points to be more specific about our research and to get a more specific result on our outcome.

- A 44 Question survey was created which included both general questions and questions related to target variables.
- Survey was distributed among individuals who were ideal for the research study.
- Only individuals who consented to provide data for the research were included. All personal information that could reveal identities was excluded.
- The research includes one major age section, which is youth and young adults. This kind of age section is from 15 years old to 30 years old, as they are considered young people. Other than this age group, no data has been collected.

- As this research consists of targeting social media influence, we considered more on the customer side and collected data about their experience and thinking.
- A questionnaire has been designed so that a definitive conclusion can be reached regarding their use of social media and their spending habits of both time and money.
- Information has been used in such a way that relations can be made between different segments of the same research work. Information that can't be used for meta-analysis or can't contribute to the whole picture wasn't considered.
- A data set has been made considering the majority group of customers and their usage patterns. Which talks about the interest of a bigger and more common population.
- Collected information that can be expressed or converted into statistics is only considered, and only information that could be quantified or converted into statistics was considered, while all other data was excluded.
- Each section has been designed to be independent and self-defining so that they do not rely on other sections. This technique and implementation contributed to the accuracy of data processing to be more accurate, and the influence of irrelevant information has been brought nearly to none.
- The research has been conducted using the English language. Due to it being the most widely spoken language, it was decided by us to be the best choice for our research. It is the international language that would also smoothen our research without causing any overcomplications.
- Collecting necessary responses turned out to be an arduous task. Due to the survey consisting of a large number of questions the data took much more time to collect than expected.
- Collected responses were then prepared to apply machine language algorithms.

3.3.2 Dataset Description

Here we have provided a detailed description of the surveyed data. A total of 1018 people participated in the survey and we have collected each and every one of their responses. A visualized representation of the statistics has also been provided.

We were able to collect data from 1018 people between the ages of 15 to 30 as this age group represents the youth and young generation. Among the 1018 people 37.9% (386) are female and 62.1% (632) are male. 8.1% are aged between 15-18 years old which is 82 people, and 276 people were within the 19-21 age group which represents 27.1%, 22-25 age group has the highest number of people, 490 people are there which is 48.1% of total numbers. Lastly, 170 people were in the group 25-30 years of age resulting in a total of 16.7%.

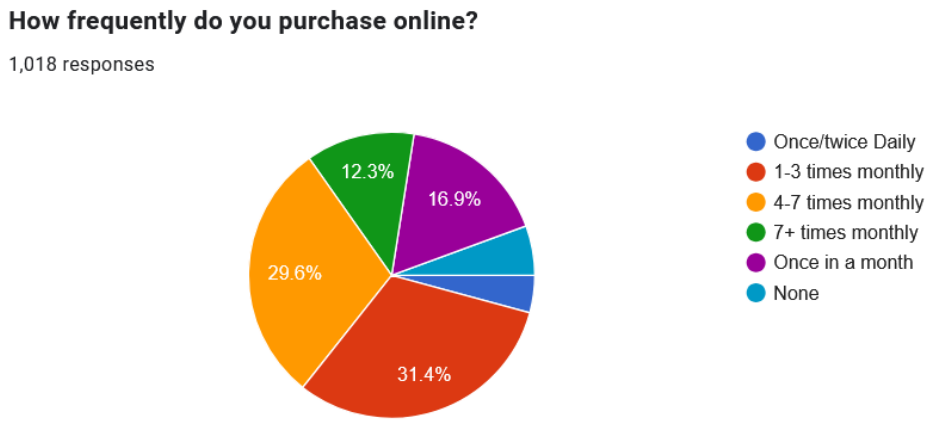

Figure 3.8: Frequency of online purchase

A total number of 43 people among 1018 purchase their necessary things online daily once/twice which is 4.2%. One to three monthly purchases have been made by 31.4% of people which is 320 people. People who do online shopping 4-7 times monthly have several 301 people (29.6%). In this era of online dependency people are more dependent on online shopping than ever, which can also be seen by the number of 125 people who do 7+ times online shopping in a month which is 12.3% of the total number. People shop from online stores once a month is 16.9% which represents 172 people. There are 57 people which is 5.6% who don't shop online at all.

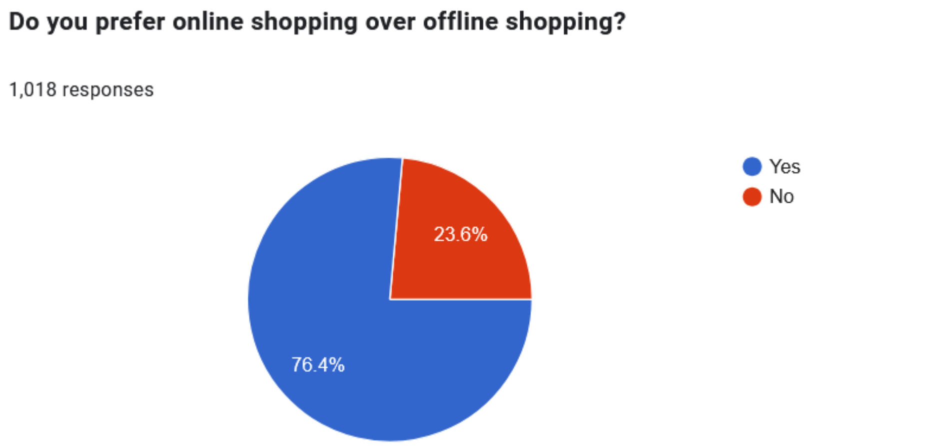

Figure 3.9: Online Shopping Recommendation

Out of 1018 people, 908 people, 89.2% people agreed on recommending online shopping to others and the rest of them said they wouldn't recommend others. 23.6% of 1018 people do not prefer online shopping which is 240 people and 76.4% people do prefer online shopping over offline shopping which is 778 people in total.

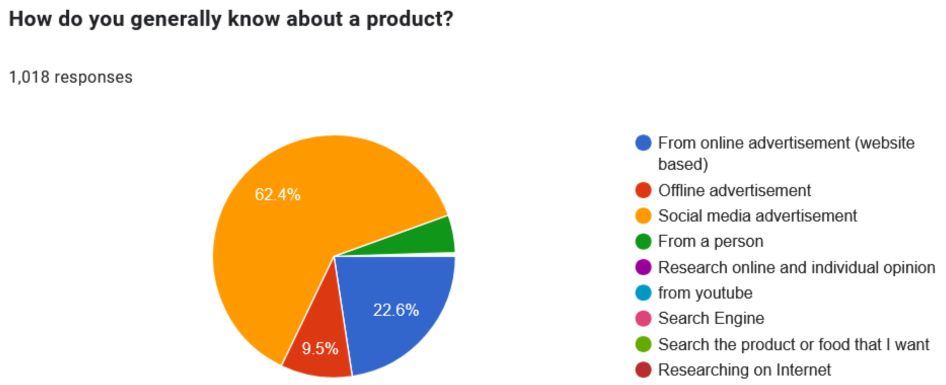

Figure 3.10: Generally learn about a Product

From the data, it was seen that about 62.4% which is 635 in number generally found out about products due to social media advertisement. About 22.6% of people which is 230 of them find out about products through online advertisements on various websites. 97 people which measures up to 9.5% of people said they usually learn about products from offline advertisements.
The maximum number of people who participated in the data collection procedure were students. 55.6% were students which is 566 in numbers. 19% are job holders and 19.4% are businessmen which equals 193 and 198 persons. 57 people are still unemployed which is 5.6% of the total number. 4 people are mixed with other jobs.

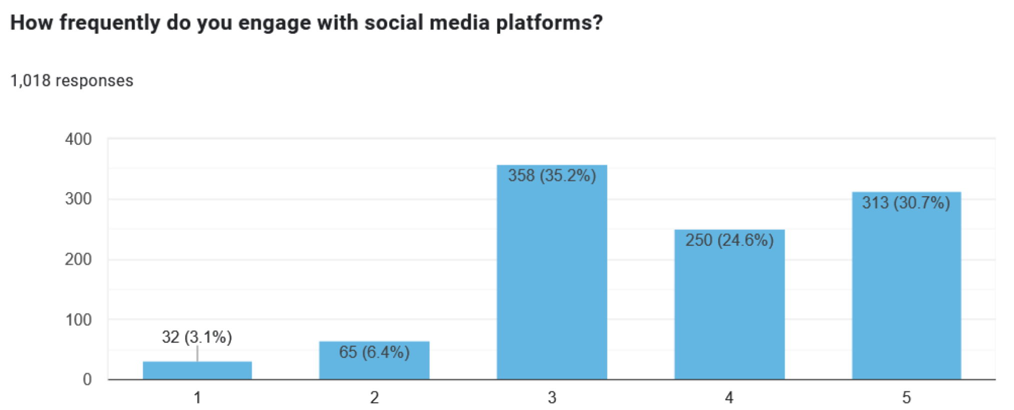

Figure 3.11: Frequency of Online Engagement with Social Media Platforms

On the frequency level of 1 to 5 where 1 represents the least frequency and 5 represents the most frequency. We got a frequency level of 5 for the people as 313 people which is 30.7% of the whole selected this option. 35.2% of people were neutral as they chose level 3. 358 people chose the neutral option. 250 people, which is 24.6% choose level 4 of frequency level. Level 2 is chosen by 65 people which is 6.4% and only 32 people (3.1%)choose level 1.

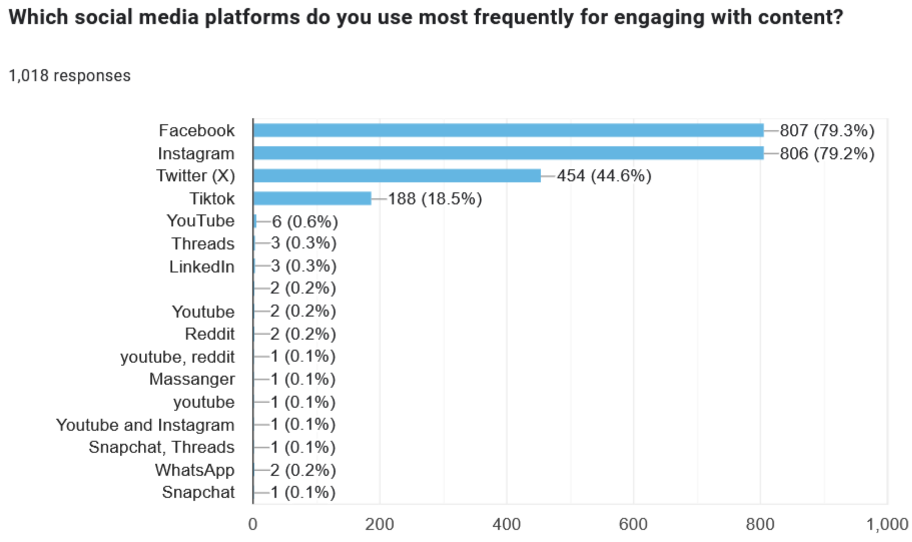

Figure 3.12: Most Frequently Used Social Media Site for Engagement

Most people use Facebook more frequently to engage with content as we got data in our support of the statement that 807 uses Facebook which is 79.3% of people. 806 people use Instagram (79.2%) and 44.6% use Twitter which is 454 people and 18.5% use TikTok and the rest of the people use other methods.

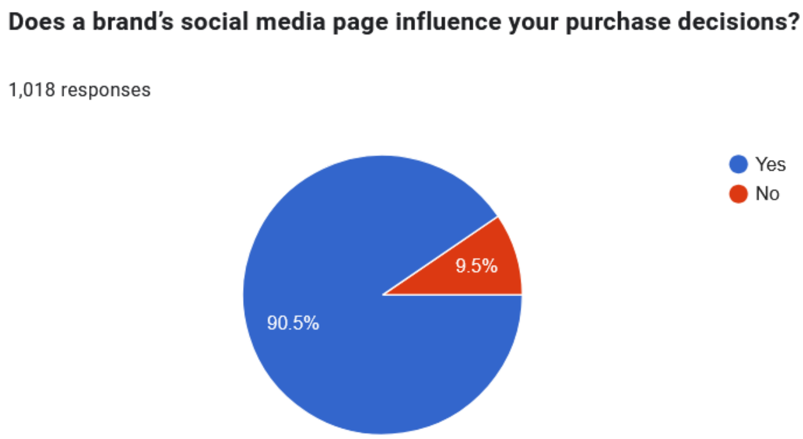

Figure 3.13: Social Media's Influence on Purchase Decision

90.5%, which is 921 people out of 1018, think that they get influenced in their purchase decision by the social media page of a brand. Only 9.5 % of people, which is 97 people in number, think the opposite. Similarly, 88.7% people said that they follow online advertised pages and 11.3% said that they don't follow.
The effect of social media sites is real as 35.3% people say that they use more than 3 social media sites. converting 35.3% into 359 people using 3+ social media sites. 34.6% people use 3 social media sites which are 352 people. 24.5% (249 people)people use 2 sites. There are only 5.7% people who use 1 site which is 58 people.
878 people said that they purchase food online which is 86.2% people. 13.8% of 140 people said that they don't purchase online food at all.

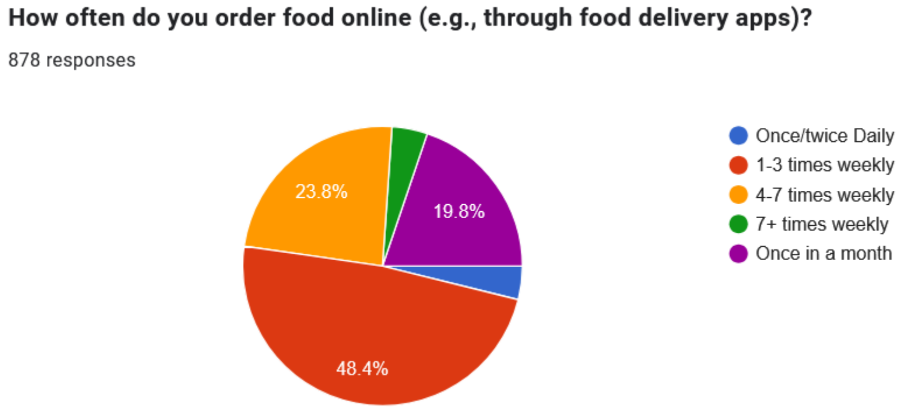

Figure 3.14: Frequency of Ordering Food Online

19.8% of people purchase food once a month out of 174 people. 48.4% of people purchase online food 1-3 times a week. 3.9% of people purchase food once/twice daily. 4-7 times weekly purchases done by 23.8% people. 4.1% of people's purchases more than 7 times weekly.
40.7% people order snacks the most and then 22.1% people order dinner. Lunch is ordered by 31.5% people and breakfast is ordered by only 5.7% people.
people mostly order and prefer fast food with 67.5% people ordering it most(combined with other food categories). Bengali food is ordered by 59.5% people combined with other food options. 56.9% of people order Thai food and 55.5% of people order Italian food. 48.2% of people order Indian food. 34.4% order Korean/Japanese foods. The rest of the people order other categories of food.

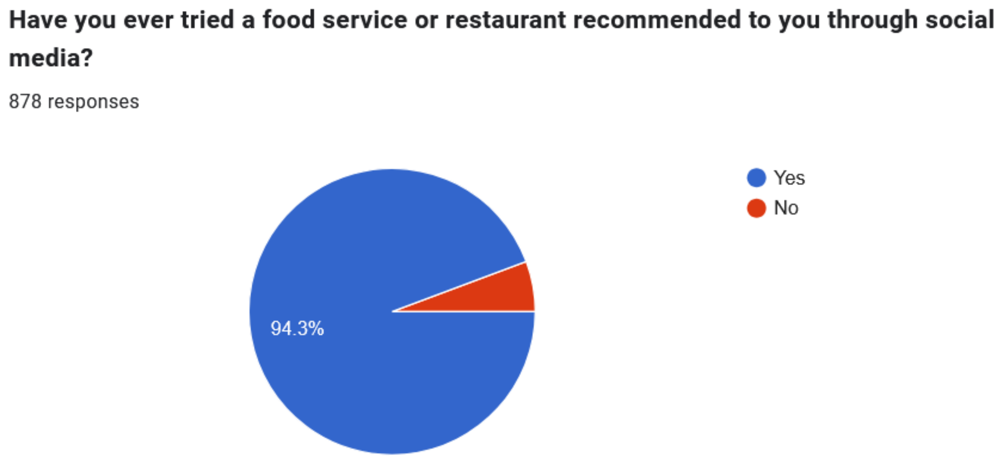

Figure 3.15: Food service/Restaurant recommended through Social Media

80.3% (705) people live with their family while others don't. 94.3% of people agreed that they tried food services or restaurants recommended to them through social media. On the other hand, 5.7% of people denied it. Delivery charge affects 91.6% people while purchasing food online and for 8.4% people, it doesn't affect. 95.6% of people said discount coupons do have an impact on choosing to order takeout and 4.4% of them said it's not that impactful.

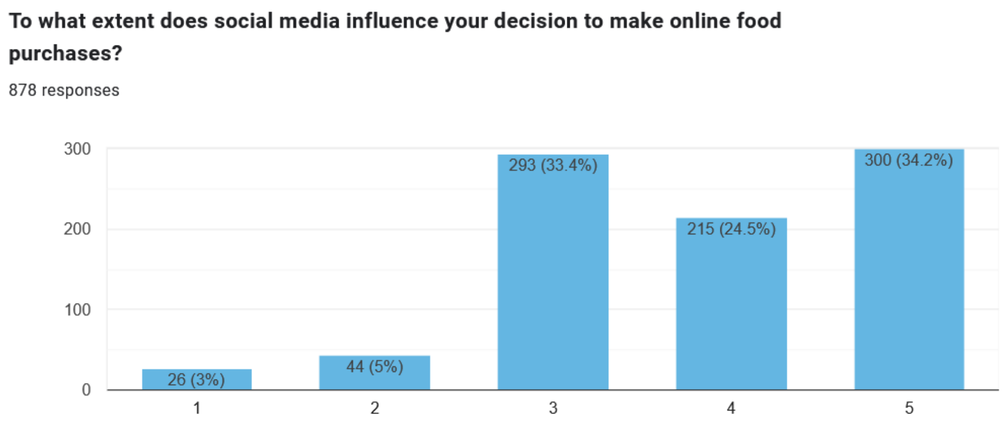

Figure 3.16: Social Media's Influence on Online Food Purchase Decision

On the frequency level of 1 to 5 where 1 represents the least frequency and 5 represents the most frequency. We got frequency levels of 5 with 300 people which is 34.2% and level 3 for 293 people which is 33.4%. 215 people, which is 24.5% choose level 4 of frequency level. Level 2 is chosen by 44 people which is 5% and 26 people (3%)choose level 1.

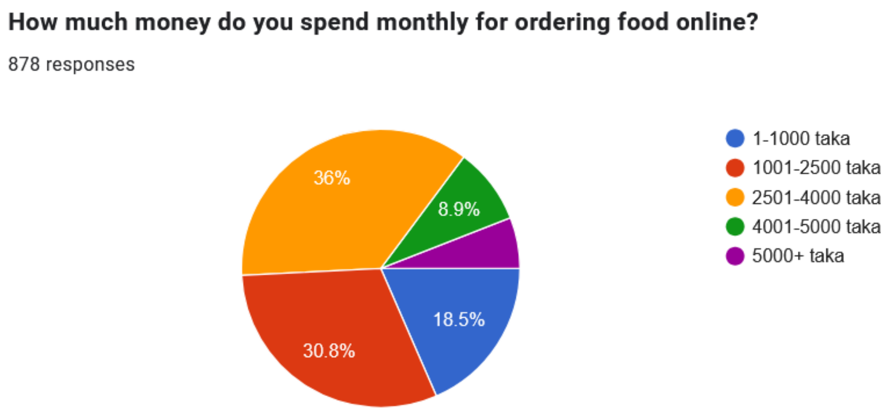

Figure 3.17: Monthly Expenditure on Online Food Purchase

18.5%(162) people spend 1-1000 taka monthly for ordering food online. 270 people, which is 30.8% spends 1001-2500 taka monthly. 2501-4000 taka is spent monthly by 36% or 316 people. 5.9% people spend 5000+ taka monthly which is 52 people. 4001-5000 taka is also spent by the 8.9% of people as last.
Most of the people prefer Foodpanda as their main food delivery platform. 588 people order through food pandas which is 67% of the total respondents.15.6% of people prefer pathao and the number is 137 of them. 75 people, which is 8.5% prefer ordering from the website of the restaurant. 47 people, which is 5.4% ordered through Hungrynaki, and the remaining 3.4% which is 30 people ordered from the pages of the restaurant.

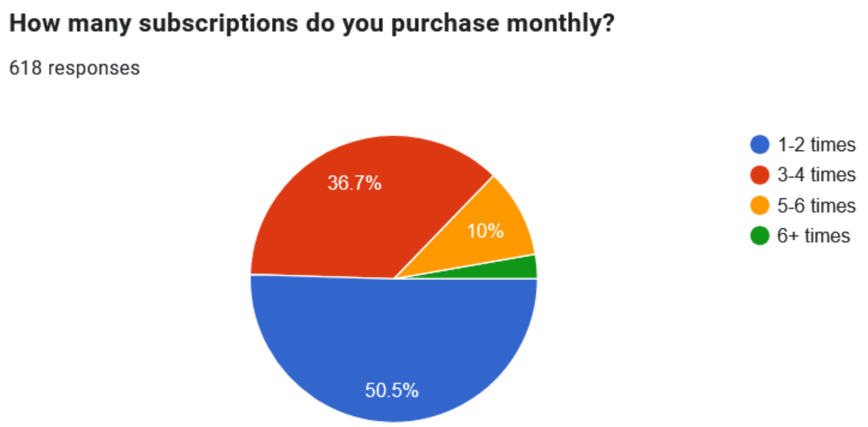

Figure 3.18: Subscriptions Purchased Monthly

About 400 (39.3%) people out of 1018 do not use any online subscription services. And 618 (60.7%) people use online subscription services.
Of that 60.7% of people, about 312 people purchase 1-2 times monthly, which is the highest (50.5%). Again, 36.7% of that (about 227 people) buy 3–4 times, and about 62 people (10%) buy 5–6 times. Moreover, 17 people buy more than 6 times monthly, which is 2.8%.

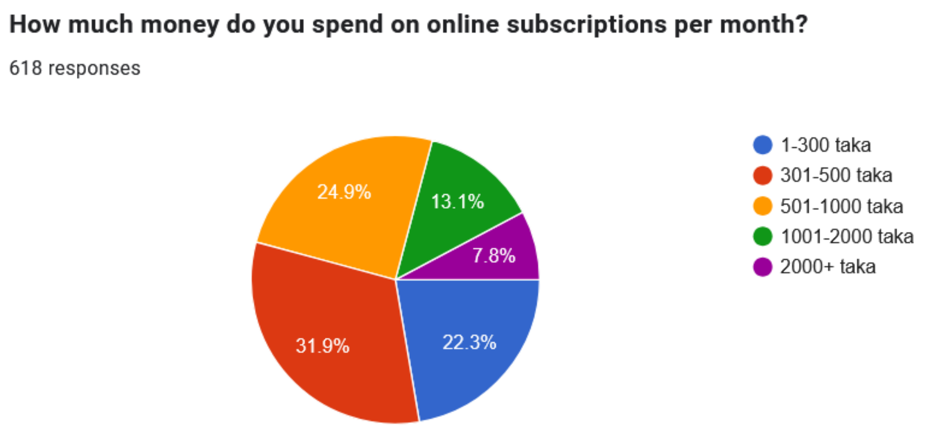

Figure 3.19: Monthly Expenditure on Online Subscription Service

In spending money monthly, 138 (22.3%) of people spend 1–300 taka. Again, 197 people (31.9%) spend 301-500 taka, 154 people (24.9%) spend 501-1000 taka, 81 people (13.1%) spend 1001-2000 taka, and about 48 people (7.8%) spend more than 2000 taka.

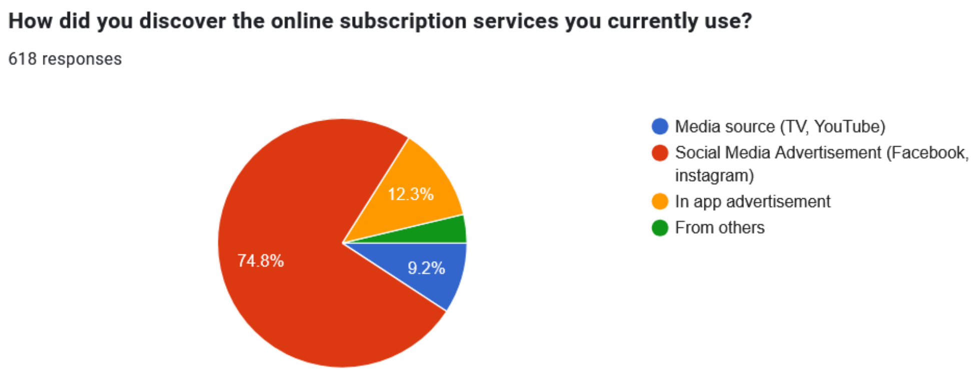

Figure 3.20: Discovering Online Subscription Service

In terms of discovering the service, about 462 people (74.8%) discovered it from social media, 57 people (9.2%) found it from media sources, 76 people (12.3%) found it from advertisements, and 23 people (3.7%) found it from others.

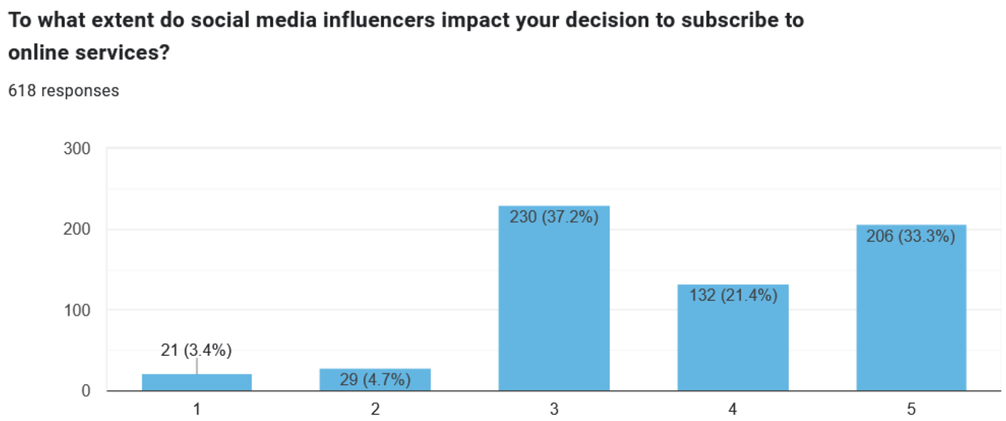

Figure 3.21: Social Media Influencers Impacting Online Subscription Purchase Decision

In terms of “To what extent do social media influencers impact your decision to subscribe to online services?” on a scale of 1 to 5, which refers to low to high, about 230 people (37.2%) selected 3, 206 people (33.3%) selected high, about 21 people (3.4%) selected low, 29 people (4.7%) selected 2, and 132 people (21.4%) selected 4. Again, in terms of payment method, 278 people (45%) selected online banking services, about 557 people (90.1%) selected mobile financial services, and about 213 people (34.5%) selected wallets or gift cards.
In terms of “Which type of subscription do you use mostly?” 139 people (22.5%) selected 1 month, about 265 people (42.9%) selected 3 months, about 154 people (24.9%) selected 6 months, and about 60 people (9.7%) selected 1 year.

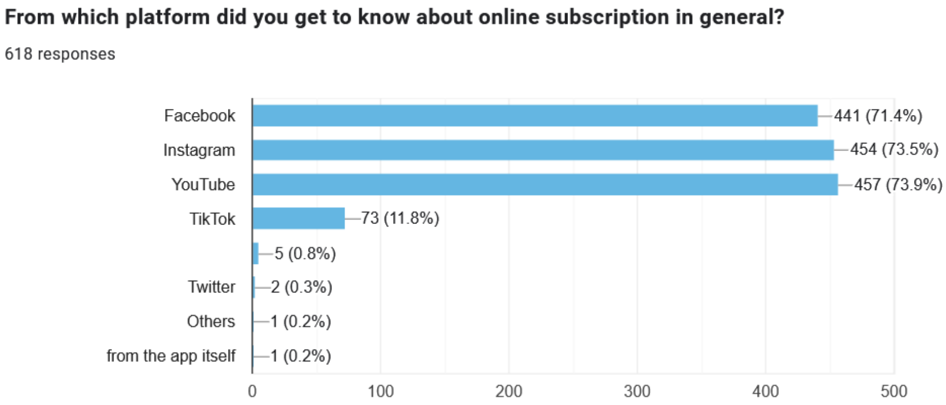

Figure 3.22: Discovered Online Subscription from which Platform

Moreover, in terms of "From which platform did you get to know about online subscriptions in general?" about 441 people (71.4%) selected Facebook, about 454 people (73.5%) selected Instagram, 457 people (73.9%) selected YouTube, 73 people (11.8%) selected TikTok, about 8 people (1.3%) selected others, and 1 person (0.2%) selected "from the app itself.".
From our collected responses of 1018 people, it was seen that about 63.9% (651) people responded positively saying they play online games. And the rest 36.1% (367) people responded negatively to playing online games.

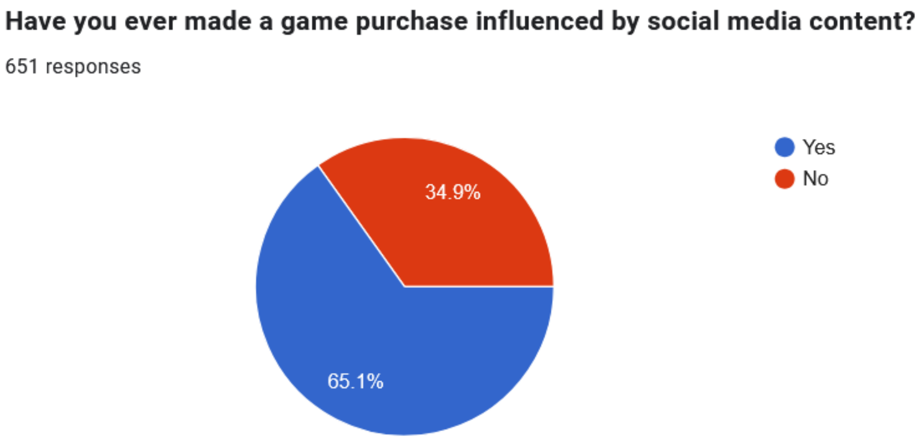

Figure 3.23: Social Media Content's Influence on Game Purchase

Data shows that 65.1% (424) of the people who have spent money on online game purchases were influenced by social media. While 34.9% (227) of people were not influenced by social media to make a purchase or didn't make a purchase at all.

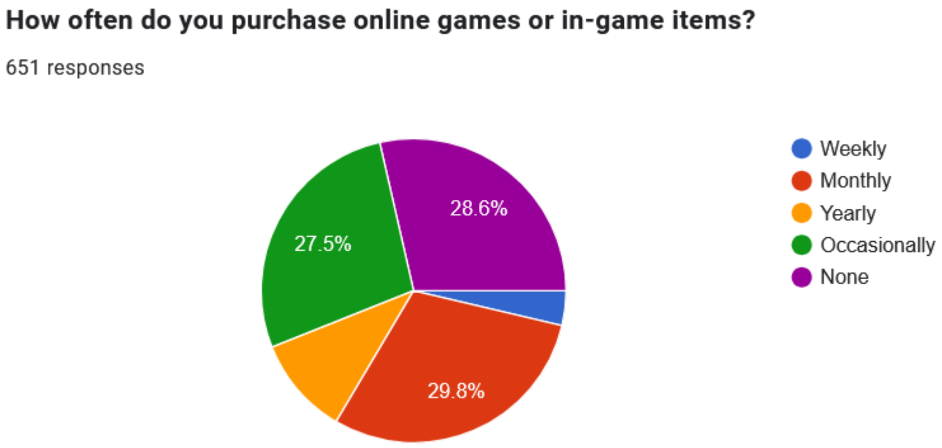

Figure 3.24: Frequency of spending Money on Online Games

In data of how often people spend money on online games, 28.6% (186) of people have never made online game purchases. 27.5% (179) of people spend money on the game occasionally. About 29.8% (194) of people purchase in-game products on a monthly basis. While 3.7% (24) of the people have shown to spend money weekly on online games and the rest of 10.4% (68) are yearly spenders.

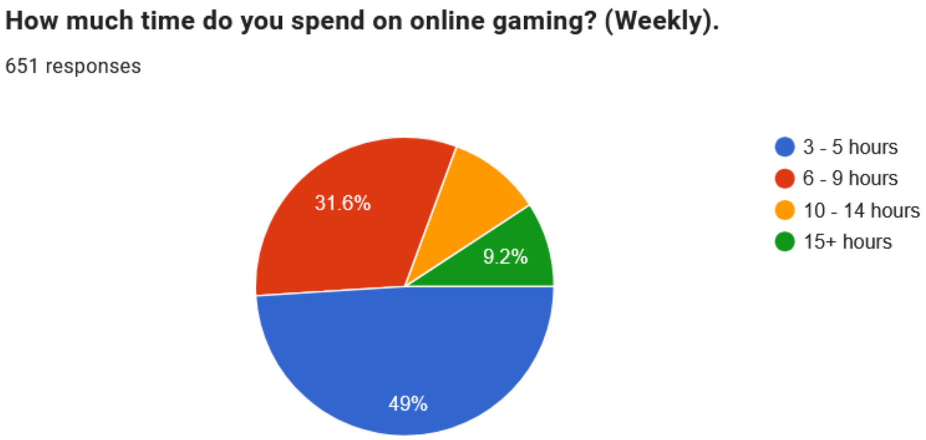

Figure 3.25: Time Spent on Online Games

In the question of how much time people spend playing online games on a weekly basis, it can be seen that 319 people, 49%, play about 3-5 hours. About 31.6% (206) of people spend about 6-9 hours playing online games. While playing for about 10-14 hours and 15+ hours have 9.2% (60) and 10.1% (66) respectively.

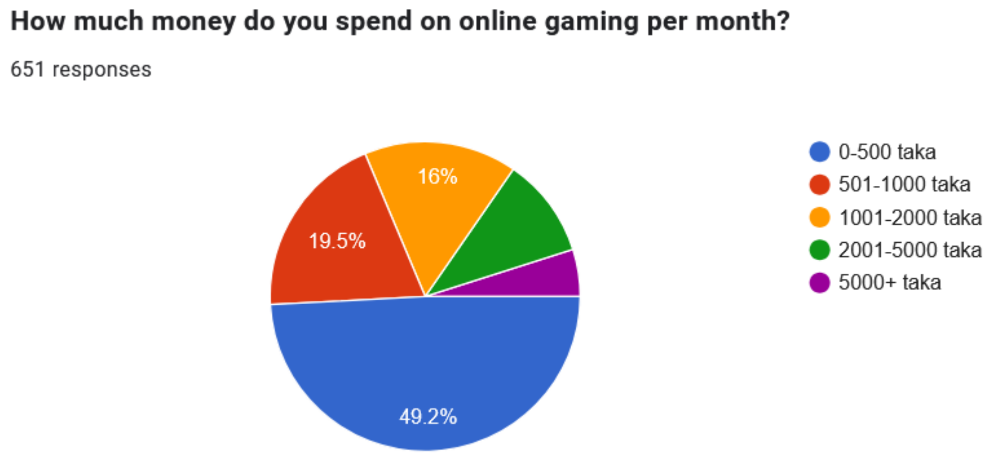

Figure 3.26: Monthly Expenditure on Online Games

From the data collected, it is seen that the amount of money that people spend also varies quite heavily. While 320 players that makeup 49.2% of the group fall in the category of spending 0-500 taka, we can see that the next biggest percentage are the people who spend about 501-1000 taka being 19.5% (127). About 16% (104) of the people spend about 1001-2000 taka and 10.4% (68) percent of people spend about 2001-5000 taka. Lastly, we see that 4.9% (32) of people are the ones who spend more than 5000 taka for their online game purchases.
The recorded data for payment methods show that 67 people of the youth group use Mobile Financial Services such as Bkash, Nagad, Rocket, etc. for their purchases regarding online games. That is about 72% (469) use Mobile Financial Services. Next, it can be seen that about 34.9% (227) of them use Online Banking Services for their expenditure. 34.4% (224) of people use Wallets or Gift Cards such as Steam, Google Play, iTunes, etc. for spending money on games. The rest of 18.9% (123) in the data do not spend money on online games.

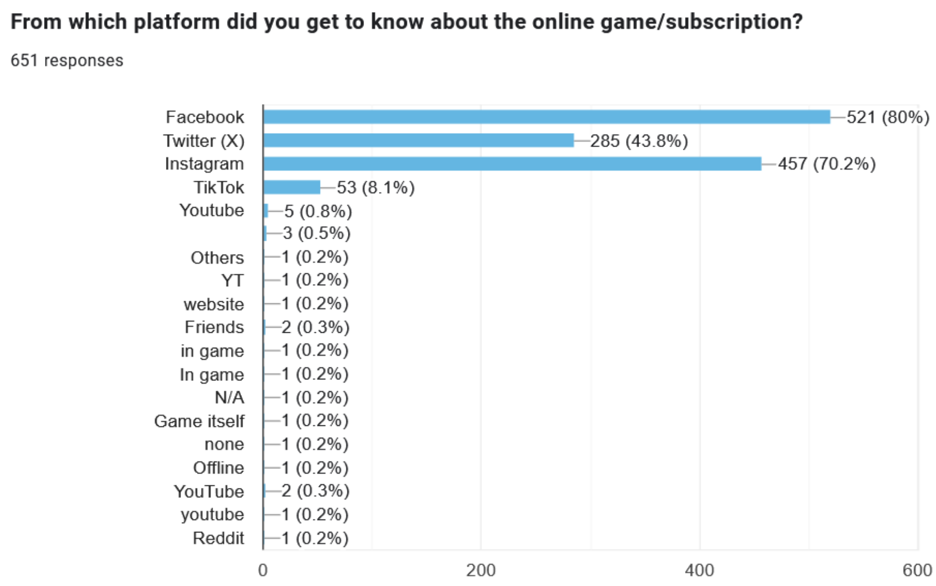

Figure 3.27: From which Platform did you learn about Online Games

The recorded data also shows that about 80% (521) of people learn about online games from the social media site Facebook. 43.8% (285) of the people find out about games from Twitter(X). And about 70.2% (457) of the people come to learn about online games through Instagram. And the remaining 4.1% selected others.

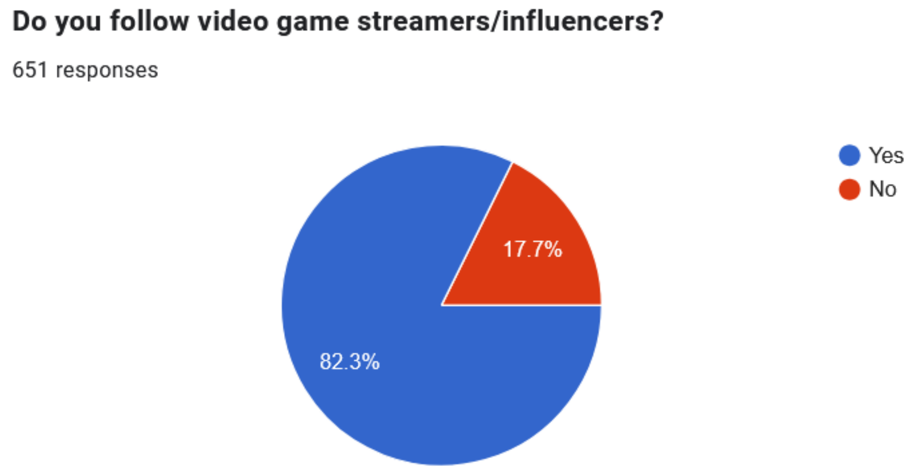

Figure 3.28: Following Online Streamers/Influencers

From 651 people who responded to the form positively about playing online games, about 82.3% (536) people said that they follow social media streamers/influencers. And the rest 17.7% (115) of people said that they didn't follow any streamers/influencers.

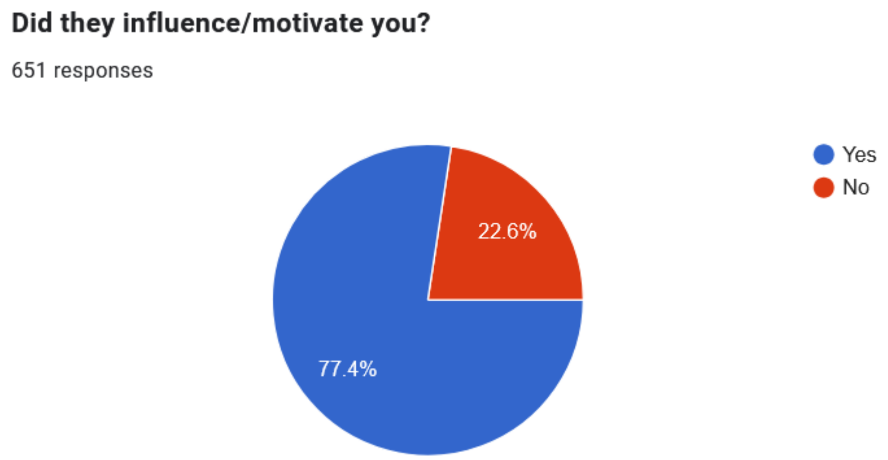

Figure 3.29: Influenced/Motivated by Online Streamer/Influencer

Again, from the same number of respondents about 77.4% (504) said that they were influenced or motivated by online creators/influencers and the remaining 22.6% (147) people said that they were not motivated by influencers.

3.3.3 Dataset Pre Processing

After collection dataset is used using CSV files. We created a separate dataset for each target variable. We then implemented this dataset into our code. CSV files were loaded using pandas and then a data frame ‘df’ was created. After that the initial data exploration is conducted using some functions like df.head, df.shape, list(df.columns), df.info and df.describe. After that, all the missing values were checked and handled. Duplicate entries were then removed using drop functions. Several columns were also discarded which we deemed unnecessary for the analysis. We then use ‘LabelEncoder’ for all categorical data and then convert them into numerical values. We then use ‘StandardScaler’ to bring all numerical values to the same scale. Correlation analysis was then computed for target variables and selected features. Our data was then split into two parts using ‘train_test_split’. Lastly, SMOTE was applied to balance the data and SimpleImputer was used to deal with any missing values.

3.4 Hypothesis Testing

Social media engagement has an impact on online food purchase, online subscription services and online game purchasing decisions. We have established some hypotheses that will explore how social media involvement affects consumers’ purchasing behavior. The hypotheses were then proved using statistical data and testing. Tests like the Chi-Square test, ANOVA, t-test, Correlation test, and Regression analysis were utilized.

3.4.1 Online Food Purchase

Social media as an indispensable part of modern life:
In recent times, consumer behaviors, which include an online shopping tendency, have changed considerably because of the growth of social media. A very notable service, which is online food purchasing, has become very popular among the youth, who are mostly active on social media and find the ordering system pretty convenient. There are different social media platforms, along with Facebook, Instagram, and YouTube, which have a major influence on purchasing behaviors, and people get to explore a wide range of food items and services through the media, as these platforms serve food advertisements, celebrity reviews, different offers, and recommendations to attract targeted people and make them order meals online. After the massive acceptance of Facebook, which was opened in 2004, consumer behavior has been massively impacted by its evolution. Like Facebook, Instagram has also been playing a prominent role in social media advertisement and attracting consumers to buy online.

The Doorstep delivery services for the time efficiency:
Social media and online services have made life easier as they are convenient and time-saving. People do not have to go outside to try restaurant food when they are living a busy schedule. They can order food through different online food delivery services and get their meals on time. Also, the contactless delivery option has become very popular, as with less interaction, users can receive the meal in place. Consumers can save their time and money and can choose their preferred item by exploring different cuisines even without tasting them by reading and watching online and social media ratings and reviews from other consumers and influencers. If those reviews were not available through social and online platforms, it could have

killed a lot of time. Thus, online services and social media engagement save more time than offline services and enhance online food purchasing decisions.

Discount offers that retain consumers:
Online food ordering services attract customers by providing their special month or occasion discounts and coupons. From the survey, we have found that 95.6% of people have an impact on discounts and coupons on their meal ordering decisions. The influencers, and food company pages promote their services through social media, and thus users get to learn about all these offers and rewards, which has an impact on their food ordering behavior.

Suitable Payment Services for Easy Purchase:
Digital payment services have been a game changer as they are more secure and fast transaction services. Social media promotes bKash, Paypal, Nagad, and other online and mobile banking services that encourage users to make purchases securely and confidentially, and users can effortlessly buy from any place without the need for hand cash. From our survey, we have seen that a higher percentage of people prefer online mobile banking as a payment option. Thus, the digital payment method has made it easier for consumers to purchase online, and we have seen from the survey that more than half of the respondents prefer online mobile banking services while they are ordering food.

H1: Consumers who have higher social media engagement spend more on online food services.
Here we will try to test our hypothesis on consumers' spending nature on online food ordering behavior that is highly dependent on their social media activity and connection by running some machine learning model that will provide the accuracy for the prediction and also using some statistical analysis that works on the categorical variables. To test our hypothesis we followed a structured process that includes collecting and analyzing data that will observe the results and that outcome will support or oppose the assumption. To examine our hypothesis Chi-square test has been used because the method analyzes the relationship between categorical variables. In this hypothesis the variable includes consumers' higher social media engagement and consumer spending nature on online food services are associated with each other. We first formulated the null hypothesis that says there is no association between social media engagement and spending behavior on online food services and then went for the alternate hypothesis which is about the association between social media engagement and spending decisions on online food delivery. For the statistical analysis, we applied the Chi-square test by relating two most relevant features, including 'How much money do you spend monthly for ordering food online?' and 'How frequently do you engage with social media platforms?'. After the testing we got the Chi-square statistic scoring 67.557 and P-value $2.662*10^{-8}$ which is significantly lower than 0.05 and rejects the null hypothesis. The rejection of the null hypothesis indicates that there is a statistically significant association between social media engagement and the consumers' spending nature on online food ordering services. The result supports the alternative hypothesis that higher social media involvement has an impact on online food purchasing. So it can be said that social platform engagement has an impact on the spending behavior of youths

while they are ordering food online. The consumers get encouragement from different celebrities and consumer reviews, recommendations, food company discounts, and rewards that they receive from social media that make them spend money on online meal services which as well as the results from statistical analysis supports the assumption that people with higher engagement in social media are those who are spending more on online food services as it is more convenient.

3.4.2 Online Subscription Services

Recommendation services that offer personalized content:
The expansion of social media has completely transformed the way people are engaged with online services. Consumers use different kinds of services, including online membership for subscription purposes, from where they can receive their preferred content according to their needs. There was a time when people had to spend a lot of time and money, and it was also hard to get their favorite trendy content on time as it was not available to receive. Now the digital era has made it easier for people to get entertained, and they can receive their favorite content on time. There are various kinds of online subscriptions available depending on different genres, including movies, music, series, telefilms, streaming services, digital publications, and others as well, and people can subscribe to these platforms. These online subscription services, including Netflix, Spotify, and Amazon Prime, offer users their preferred and personalized content based on their likings and interactions. By using some algorithms, these sites give movie suggestions, music playlists, and other services according to customer needs. Thus, the subscription services use recommendations by using some techniques that offer the best kind of content to the individual customer and convince them to buy the services.

Access to thousands of on-demand content in the subscription platform:
Online subscription services such as Netflix, Amazon Prime, and Spotify provide on-demand media services that allow the user to watch movies, and TV shows, and listen to music according to their mood and time. This convenient service has opened access to the vast media sector. They can enjoy thousands of content and options through the blessing of these services. Online subscription-based services have replaced the old system of media-consuming behavior and made it easier to receive entertainment. Before these services were available, people had to wait for timed TV programs, CD services, and music albums, but now they can enjoy the contents anytime they want. Young users are highly influenced to buy these subscriptions through the contribution of social media such as YouTube, Facebook, Instagram, and TikTok, as these apps spread the content via advertisements, celebrity posts, and videos. These social media sites are the major tools because they serve the latest movie trailers, behind-the-scenes music, videos, and amazing offers for promotion purposes of particular subscription services.

Flexible Membership Plan that attracts consumers to buy the subscription:
Different subscription services provide flexible purchasing plans like monthly, quarterly, and yearly payment methods, and a buyer can choose specific plans according

to their budget. Also, sometimes consumers can purchase a group plan through which multiple users can share the subscription plan. The membership plan offers various rewards and additional benefits to the user of specific subscription services. This way, more people are influenced to use subscription services.

H2: Social media influencers have a significant impact on the decision-making process of young users.

In this report, we will try to evaluate the hypothesis that the relationship between social media engagement and subscription exists by applying some machine learning models that will predict the behaviors of the consumer. Also by performing the Chi-square test, we will prove the hypothesis. First, we formulated the null hypothesis that shows there is no significant association between social media influencers' impact and how they influence young users to buy subscriptions. To prove our hypothesis by Chi-square test we tested with two related features which are 'To what extent do social media influencers impact your decision to subscribe to online services?' and 'How many subscriptions do you purchase monthly?'. After testing we obtained the results of Chi-square statistic 116.218 and P-value 3.5*10^-19 which is below the commonly used level 0.05 that rejects the null hypothesis. This means there is statistically a significant association between social media engagement with the impact of influencers and their influence on users' purchasing decisions. which supports the alternative hypothesis. The young consumers who spend more time on social media get more in touch with the influencers who affect their decisions and encourage them to buy. Through the influencer's recommendations, created content users get motivated to spend on online subscriptions which validates the hypothesis that social media plays an important role in youths' decisions to purchase online subscriptions. Consumers who are highly active on social media are the main users of online subscription services, as they get highly inspired by social media influencers to make subscription purchases also the statistical analysis supports our hypothesis that the buying behavior of online subscription services is connected to online social media influencers who have a huge impact on the young user purchasing decisions.

3.4.3 Online Gaming

Digital purchase or Live service games are the future of gaming:
The past two decades show how far online gaming has come as a youth's hobby. They are getting influenced by social media platforms to spend more time and money on online gaming from the promotion videos and game developers engaging management. Facebook, Instagram, and YouTube have become the biggest marketplace for video gaming hubs, as it is convenient for game developers and companies to use social media as a tool for advertisements and streamers through which they can connect to their target audience. Consumers get suggestions based on their gaming tastes, real-time content, and direct access to influencers with the help of social media, which increases user engagement. It was seen in our collected data that 63.9% of the people partake in online games. And out of those people, 65.1% have spent money on these online games due to the influence of social media.

In-game purchases are the most effective method to make players spend: Youth who follow streamers, take part in live broadcasts, and interact with gaming communities on social media get more attracted to purchasing online games. It is easier for gamers to look for the latest games and in-game purchases since they are constantly receiving gaming information, reviews, exclusive deals, and promotional offers through media platforms. Our survey shows that only 28.6% of people have never purchased any online games or in-game items. Thus, making the remaining 71.4% of people who regularly or semi-regularly make online game-related purchases.

Streamers/Influencers the idols of modern youth:
The youth are exposed to different types of gaming advertisements, streamers, and game-related content through social media as they are connected to the contacts of game developers, influencers, and companies, which enhance their purchasing decisions. From our surveyed data, we have seen people spend money on online gaming. It shows that about 82.3% of people who participated in the survey follow or watch video game streamers or influencers. And it leads to them being idolized by the youth, which motivates them to spend money on the games their favorite streamer/influencer plays.

H3: The youth who purchase online gaming services or spend money on games are overwhelmingly influenced by social media for their decision to make a purchase.

Here we will try to evaluate the hypothesis of buying online game services through the influences of social media, and with the help of different machine learning models, we will predict the relationship between online game purchasing behavior and social media connectivity. First, the hypothesis has been proved by the Chi-square test as it gives strong evidence by supporting the relationship between social media engagement and its impact on online game purchasing decisions. We applied two features, including 'Have you ever made a game purchase influenced by social media content?' and 'How frequently do you engage with social media platforms?' for the Chi-square statistic and the statistic provides a significance of 12.454 and a P-value of 0.0143, which is below the standard count of 0.05, and rejects the null hypothesis which is that there is no association between social media involvement and game purchasing decision influenced by social platforms. This statistic supports the alternative hypothesis that says there is an association between the frequency of social media engagement and game-purchasing behavior. The lower P-value indicates that the observed correlation happened because of user involvement on social media that has influenced game purchasing decisions. Game streamers and game developers in social media provide reviews, recommendations, and promotions that attract the gamer to buy games which validate the hypothesis that social media plays an important role in youths' decisions to make online game purchases. So we can say that consumers who are actively using social media are the main buyers of online gaming services. This data and statistical result pretty much support our hypothesis on the purchasing behavior of online gaming services, which is directly connected to online social media engagement.

3.4.4 Relation between online food purchase, online Subscription service purchases, online game purchases

The internet created many opportunities and services both for the consumer and provider. It is like an ocean of resources. One resource or opportunity opens the door to others. So, it makes it easier to assume that if a person is having one online service then he also may be on the more inclined side to using other services as well. For example: if a person makes online food purchases then there are chances that he will be into purchasing subscription services online or purchasing games online as he is already familiar with the online environment and it is way more convenient for him.

H4: A person interested in online food purchases is also likely to be interested in purchasing online subscription services or spending money on online games.

Today's youth are heavily engaged in online spaces. Day by day we see more young people get addicted to the internet. This in turn increases people's demand to have more and more services available to them through online means due to it being more comfortable. From our survey, we saw that 76.4% of the people preferred online shopping over offline. This shows that people prefer to use online shopping services more and can be applied to all of our target variables. Similarly, if a person tends to buy more products online then there is a high possibility for them to buy a different kind of product also through online shopping. To prove our point we have conducted correlation tests and Chi Square tests. We selected features that show the relationship between online food purchases, online subscription services, and online gaming. An independent variable and 2 dependent variables have been selected to see if they share a strong association. The Independent variable is 'Do you use online food purchases? ' And two dependent variables are 'Do you use any online subscription services (e.g., streaming services, digital publications)? ' and 'Do you play online games? '. We obtained two different Chi-square values and p-values for two variables. For the first one, the Chi value is 41.30125 and for the second variable, the Chi value is 8.113425. These values indicate the strength of the association between the independent variable and the dependent variable. With these values, we can understand the difference in the strength of the relation between variables. Moreover, we got two different p-values for two dependent variables. For the first variable, the value is 1.3048567*10̂(-10) and for the second variable, the value is 0.0043939. Here the p-values are less than 0.05. If the p-value is less than 0.05 then the variables have a strong relationship. So, it can be said that the variables are strongly connected. Thus, we can prove that if a person is interested in purchasing food online, then he/she will also be interested in purchasing subscriptions or online games

3.4.5 Null hypothesis

Frequency of social media usage is incredibly high:
We researched how frequently these people engage with social media platforms on a scale of 1 to 5 (lower to higher). 55.3% of people replied with 4 or higher, while

35.2% stayed in the middle with 3, and only 9.5% of people selected 2 or lower. We also see how the majority of them use social media sites like Facebook, Instagram, Twitter (X), etc. This shows us that more than 90.5% of people use social media on an average to heavy frequency level.

Social media pages swaying purchase decisions:
Without much surprise, it was also noted that 90.5% also answered “yes” when asked if a brand’s social media page influenced their decision to purchase or not. 88.7% also replied with “yes” when asked if they follow online advertised pages on social media.

Social media platforms and influencers on the platform affect people’s opinions:
For example, people who play online games gave a positive reply, with 65.1% of people saying they were influenced by social media content to make in-game purchases. They also responded with 77.4% saying how online influencers and streamers influence them or motivate them.
NH1: There is no significant positive correlation between the Frequency of social media usage and how it affects the online shopping preferences of youth and young adults.

Addressing the claim that there is no positive correlation between the frequency of social media usage and how it affects people’s shopping preferences. This claim can be refuted. Routing back to the information presented, we can see how youth and young adults prefer social media advertisements and recommendations more than other options available. This was also again seen when the research focused on the target variables, such as online food purchases, online subscriptions, and online gaming purchases. Applying the Chi-square Test between relevant features a result of 115.4881214749985 which is relatively high. A high Chi-square score denotes that the features used are not independent. The two most relevant features used for the test were ’How frequently do you engage with social media platforms?’ and ’Does a brand’s social media page influence your purchase decisions?’. Next, the p-value was found to be 4.90949727431894*10^-24 which is much lower than 0.05. A p-value lower than 0.05 signifies the existence of a strong association. So, the testing rejects the null hypothesis and establishes a strong association between the features. Thus, it can be seen how most youth and young adults nowadays frequently engage with social media platforms and how the frequency of their engagement has influenced their purchase decisions. Therefore, it is very well proven that there is a significant positive correlation between the frequency of social media usage and people’s online shopping preferences which in turn makes the first claim of the null hypothesis refutable with relevant evidence.
Modern-day persuasion to sell products is Online advertising:
From our collected data of 1018 individuals, the responses show that 62.4% of the respondents generally find out about products from social media websites. And 76.4% prefer shopping online over offline. While 89.2% of the respondents agreed upon recommending online shopping to others.

The big three of social media sites:

Our survey also shows that the respondents' most used social media site is Facebook, with 79.3% of people who use it regularly. And with a very minute difference, Instagram is just behind it with 79.2%. Third place is Twitter (X), which still has a relevant number of users with 44.6% of people.

Social media usage creates more online discussions:
We researched to what extent people participate in discussions or activities related to online subscriptions, game purchases, or food orders on social media on a scale of 1 to 5 (lower to higher). And 51.2% of people replied with 4 or higher, while 33.2% stayed in the middle with 3, and only 15.6% of people selected 2 or lower.

NH2: The type of social media platforms such as Facebook, Instagram, and Twitter don't have significantly more influence over online shopping preferences than other online sites or shops.

For the second claim, through our conducted survey and data analysis, we have found that people are more likely to purchase online due to social media advertising. Applying the Chi-square Test between relevant features a result of 183.52710820880907 which is relatively high. A high Chi-square score denotes that the features used are not independent. The two most relevant features used for the test were 'Which social media platforms do you use most frequently for engaging with content?' and 'Does a brand's social media page influence your purchase decisions?'. Next, the p-value was found to be 3.528517814819318*10^-23 which is much lower than 0.05. A p-value lower than 0.05 signifies the existence of a strong association. So, the testing rejects the null hypothesis and establishes a strong association between the features. Thus, from the research we conducted using 1018 youths and young adults as participants, we can conclude that the type of social media that people use, i.e., Facebook, Instagram, Twitter (X), etc., has significantly more influence on them compared to normal or even online advertising, which isn't done through social media. This results in people preferring to view advertisements and make purchases through social media more than online shops or sites. So, the second claim of the null hypothesis can now be clearly disproven using our research as evidence.

In this era of social media, consumers highly depend on what is available on the internet. They like to do research about their product and what they are getting. They also compare a product with another and then conclude. They harness the knowledge that is available on the internet before purchasing and also search for other customers' experiences who have already undergone the process of using a similar product. These benefits of the internet have made them more addicted to online shopping, and this craze of online shopping is hugely influenced by social media platforms. In every scroll or moving to the next page, social media platforms either advertise a product or someone from social media posts something that attracts others to follow and like that product. These statistics show that if algorithms and methods are implemented correctly, they can show more specific suggestions to those who want to purchase or are confused about whether to purchase or not. Also, social media spending can be saved by studying about spending and limiting it to the necessary things only. A study can also be conducted to know about consumer likings and behaviors, which can be used for the betterment of brands to serve cus-

tomers more effectively. Also, market studies will help newcomers establish their businesses targeting a specific portion of consumers.
As the modern youth are heavily engaged in social media platforms and follow online creators/influencers, it is advised to use these creators/influencers to create more awareness and engagement for products through advertising. This meta-analysis will give insight to people in the business sector in deciding how to approach modern-day youth and young adults and capture their attention through social media platforms and influencers present on said platforms.

3.5 ML Model Preparation and Deployment

We have used the Python programming language. So, we had to use libraries like numpy, sklearn, panda, etc. We made a CSV file of the dataset with the collected data. Then we put our dataset in Google Drive and mounted it by the drive link with the code. Then we used the code to run six machine learning algorithms: Random Forest, Naive Bayes, Decision Tree, Gradient Boost, XGBoost, and Ensemble Model.

3.5.1 Online Food Purchase

The code used implements machine learning algorithms to predict if people use online food purchases based on their social media engagement. Firstly, we started our code by importing various Python libraries. The Python libraries that we imported are Pandas, Sklearn, Numpy, Seaborn, google.colab and matplotlib. Then we converted all surveyed data into a CSV file, allowing us to implement it to the required code. Next, we import the CSV file into the code by mounting Google Drive and accessing it. For this, we used "drive.mount('/content/drive')" and then"pd.read_csv" to load the CSV file. Then we stored the dataset in variable "data1". The code is then used to analyze the structure and content of the dataset. Thus, a data frame is created from the dataset. Next, the missing values are checked and handled using"check_missing_values(df)" and "handle_missing_values(df)". We then preprocess the data by selecting the features from our collected dataset. For the preprocessing, we have implemented various steps, including data dropping. We dropped unnecessary columns from the data frame. Next, we encoded categorical values using "LabelEncoder". After that, various feature selection techniques were applied. Correlation analysis is then performed to find the correlation between features and target variables. A correlation heatmap is then generated to visualize the relationships between features. Thus we removed all other features from the dataset that did not correlate with our target variable of online food purchase. Next, we split the preprocessed dataset into training and testing sets. To balance the data, we have used SMOTE because the majority class has a significantly larger number of instances than the minority class. We have also used K-fold cross-validation to ensure performance assessment.
Next, we applied the machine learning algorithms to our dataset. These algorithms include Random Forest Classifiers, Naive Bayes, Decision Tree Classifiers, Gradient Boost Classifiers, XGBoost, and Ensemble Models. Then we evaluated the performance of the algorithms. For each model, we evaluated their accuracy, classification report, confusion matrix, recall, heatmap, and cross-validation score. Then we used Matplotlib and Seaborn to help us visualize the results. We visualized the confusion

matrices, classification reports, and feature importances. Thus, we acquired a comprehensive report on the machine learning algorithms determining whether a person who has high social media engagement uses online food purchases or not.

3.5.2 Online Subscription Service

This code has been implemented to predict people's online subscription service purchase decisions using machine learning algorithms. Similar to our coding for the online food purchase target variable, we begin our code again by importing various Python libraries. We imported the following Python libraries: matplotlib, google.colab, sklearn, numpy, pandas, XGBoost and seaborn. Next, to implement the surveyed data into the necessary code, we converted it all into a CSV file. Then we imported the CSV file into the code we intended to utilize. To load the CSV file for this, we used the "pd.read_csv" technique. The dataset was then saved by us in a variable named "data1." After that, we chose the features from the dataset we've gathered and started the preprocessing of data. We used several preprocessing techniques, such as data removal and renaming. We removed columns from the data that were not necessary. Next, we used "StandardScaler" and "MinMaxScaler" to scale numerical variables and "LabelEncoder" for encoding categorical data. After that, various feature selection techniques were then used. Recursive Feature Elimination (RFE) was applied to identify the features that are essential to the target variable of online subscription services. The top 20 features of said target variable were then selected using (RFE). As a result, we were able to eliminate other features from the dataset that had no important correlation to our target variable, which was online subscription services. The preprocessed dataset was then subsequently split into training and testing sets. Due to there being far more instances of the majority class than the minority class, we utilized SMOTE to balance the data. For XGBoost we used "scale_pos_weight" parameter for handling imbalance. We also used the Ensemble model by combining a Voting Classifier, combining Random Forest, Logistic Regression, and XGBoost.
Next, we used our dataset to apply the machine learning algorithms. These algorithms include the Gradient Boost Classifier, Decision Tree Classifier, Random Forest Classifier, XGBoost, Naive Bayes, and Ensemble Model. Next, we assessed the algorithms' performance. We then assessed the accuracy, k-fold cross-validation score, confusion matrix, recall, and classification report for each model. Furthermore, to assist us in visualizing the outcomes, we utilized Seaborn and Matplotlib in the code. The feature importances, classification reports, heatmaps, scatter plots, and confusion matrices were all visualized using the above 2 libraries. As a result, we were able to obtain an extensive report on the machine learning algorithms that are utilized to ascertain whether or not an individual with a high level of social media activity is more likely to purchase online subscription services.

3.5.3 Online Gaming

Following the same steps as our previous two target variables, we started the coding part for online gaming. Data was taken from the CSV file of the survey that we conducted. Next, the Python libraries were implemented. The Python libraries used for the code were pandas, sklearn.preprocessing, numpy, sklearn.impute, seaborn, google.colab and matplotlib. Then the dataset was loaded using pandas' "read_csv" function. Functions such as df.head, df.shape, df.info, and df.describe were then used to initially explore the dataset. We dropped unnecessary columns from the data frame. Next, we encoded categorical values using "LabelEncoder". The dataset was then saved, and we proceeded to select only the features that were relevant to our target variable of online gaming. Recursive Feature Elimination (RFE) was applied to identify the features that are essential to the target variable of online subscription services. The top 20 features of said target variable were then selected using (RFE). Exploratory Data Analysis (EDA) is applied next. We created bar plots using the function "sns.barplot" for visualizing along with a correlation heatmap. After the dataset was split into two parts, one was the training set, and the other was the testing set. We then applied SMOTE for balancing. K-Fold Cross-Validation has also been used for model evaluation.

Finally, we applied the machine learning algorithms to our dataset, similar to how it was done for previous codes. The algorithms were Random Forest Classifier, Naive Bayes, Decision Tree Classifier, Gradient Boost Classifier, XGBoost, and Ensemble Model. A loop was used to test all models. Next comes the evaluation of the performance of the algorithms. Every model was then evaluated, checking their accuracy, classification report, recall, and k-fold cross-validation score. Lastly came the visualization part. We visualized the heatmap, classification reports, accuracy, and feature importance. So after going through the outcome, we can believe that the influence of social media is very prominent when it comes to making purchases regarding online games.

Chapter 4

Result and Discussion

4.1 Result Analysis

We have got a variety of results from our models. We have 3 sections due to our target variables. Thus, for each section, we have run six different algorithms. The results we attained each differ from the other. We have measured the Accuracy, Precision, Recall, and F1 score for these models.

4.1.1 Result Online Food Purchase

Table 4.1: Accuracy, Precision, Recall and F1-score of online food purchase for the 6 Different ML Models

Model Name	Orders Food Online				Doesn't Order Food Online			
	Accuracy	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	0.92	0.93	0.98	0.95	0.92	0.81	0.51	0.63
Naive Bayes	0.79	1.00	0.75	0.86	0.79	0.39	1.00	0.56
Decision Tree	0.89	0.93	0.94	0.94	0.89	0.59	0.58	0.58
Gradient Boosting	0.92	0.95	0.95	0.95	0.92	0.69	0.68	0.69
XGBoost	0.90	0.93	0.95	0.94	0.90	0.64	0.56	0.60
Ensemble Model	0.92	0.94	0.98	0.96	0.92	0.80	0.58	0.67

Initially, the step was taken to study the dataset and about 1018 people were surveyed who were mostly teenagers and young adults, and they responded to questions about their social media usage and online meal ordering behaviors to develop the dataset. Then we applied the collected dataset to our code to achieve the result. To train and test our dataset, we worked on specific features by using machine learning language, and among them, the most important features were the frequency of the uses of social media, consumers' monthly spend on total food orders that were placed online, their most used social platforms and influence of recommendations and advertisements on food on their purchase decisions. We have relied on some machine learning algorithms, including Random Forest, Decision Trees, Gradient Boosting, Naive Bayes, XGBoost, and Ensemble Model to evaluate the precision of hypotheses. After using smote and K-Fold the outcome of the Random Forest model was great, as it produced the most accurate result by scoring 92% in predicting the engagement of users with social media to order food online. For class 0, the precision report was 0.81 recall 0.51 and the f1 score was 0.63. But for class 1 the score was high and it provided a precision report of 0.93, recall 0.98, and f1 score of 0.95 which denotes that the results accurately assume every single scenario Compared to other models, the performance of Naive Bayes sometimes struggled to classify the data correctly and gave a poor result and here it achieved 79% accuracy, and again for class 1 the precision was 1.00, recall 0.75, and f1 score 0.86 and for class 0 the precision report 0.39, recall 1.00, and f1 score was 0.56. The Decision trees, which can work with complex datasets, usually give high precision and in this case, it also accomplished 89% accuracy for the model the precision report was 0.59, recall was 0.58 f1 score was 0.58 for 0 class and the precision report was 0.93; recall 0.94 f1 score was 0.94 for class 1. Similarly, Gradient boost being a robust model as well, achieved the results of 92% accuracy in predicting with a precision of 0.69, recall of 0.68, and f1 score of 0.69 for 0 class and for class 1 the precision report of 0.95, recall of 0.95, and f1 score 0.95. We also applied XGBoost and got an accuracy of 90% for class 1 the precision was 0.93, the recall was 0.95, and the f1 score was 0.94 and for class 0 the precision was 0.64, the recall was 0.56, and the f1 score was 0.60. Lastly, we applied the Ensemble Model and achieved 92% accuracy with a precision of 0.80, recall of 0.58, and f1 score of 0.67 for class 1 got a higher precision of 0.94, recall of 0.98, and an f1 score of 0.96 which denotes a higher social media engagement. The performance of these models was assessed by using the k-fold cross-validation procedure based on recall, accuracy, precision, f1-score, and the confusion matrix to determine the most effective model for predicting online food order behavior based on social media engagement.

These models provide high accuracy, which indicates that social media engagement is the most effective medium for consumers for their online food ordering decisions. They can easily discover new food items by watching promotions, recommendations, and reviews on the pages of food service companies through social media like Facebook, Instagram, and YouTube. Young adults who are highly active on social media use the advantages effortlessly and get the easiest access to these food services. So 94.3% of consumers who tried food through social media recommendations and their dependency on discounts and offers to make decisions on buying online food are the people who spend money monthly on online food services, and the machine learning algorithms that were used for prediction provide high accuracy as social media directly affect the consumers purchasing decision, supporting our hypothesis that online food ordering behaviors and the amount they spend monthly on online food ordering services get impacted by high social platform engagement.

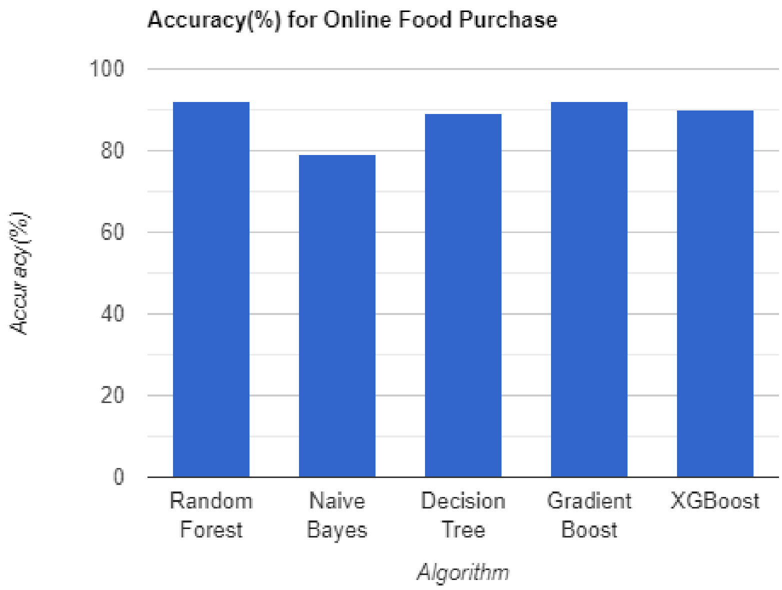

Figure 4.1: Accuracy Graph for Online Food Purchase

4.1.2 Result Online Subscription Service

Table 4.2: Accuracy, Precision, Recall and F1-score of Buys Subscription and doesn't Buy Subscription for the 6 Different ML Models

Model Name	Buys Subscription Online				Doesn't Buy Subscription Online			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Random Forest	0.67	0.71	0.78	0.74	0.67	0.59	0.50	0.54
Naive Bayes	0.68	0.69	0.86	0.77	0.68	0.66	0.41	0.50
Decision Tree	0.60	0.67	0.68	0.67	0.60	0.49	0.48	0.48
Gradient Boosting	0.67	0.70	0.82	0.75	0.67	0.61	0.45	0.52
XGBoost	0.65	0.70	0.75	0.72	0.65	0.56	0.49	0.52
Ensemble Model	0.68	0.72	0.78	0.75	0.68	0.60	0.52	0.56

The first step in our study was to form a dataset using our collected responses from 1018 respondents. Then the data was used in our code to achieve the required results. The data was then trained and tested to acquire the following results: Using machine language, we selected specific features that were relevant to our target variable. After that, the code determined and sorted the features according to their importance. It was mostly focused on some particular features that include the users' social media engagement, the way they have discovered their currently used subscription through social platforms, the impact of social media influencers, reviews, recommendations, and advertisements on their purchasing decisions and about the social media platform they used to learn about the preferred subscription service. To examine how social media engagement has impacted online subscription buying behaviors, we have applied some algorithms, including Random forest, Decision trees, Gradient boost, Naive Bayes XGBoost, and Ensemble Model that were also used for our other target variables. After applying smote and K-Fold the Random Forest model performed moderately, accomplishing 67% accuracy in

its prediction of users' involvement in social media to buy online subscriptions and generating a decent outcome with the f1 score 0.54, recall 0.50, and precision reports 0.59 for 0 class and for class 1, the precision report was 0.71, recall 0.78 and f1 score was 0.74 on decision-making purposes of online subscription. Next, looking at Naive Bayes, it was seen that it produced an accuracy of 68% in prediction. The precision was 0.66; recall 0.41 f1 score was 0.51 for 0 class and the precision report was 0.69, recall 0.86, f1 score was 0.77 for 1 class. The Decision Tree produced predictions with 60% accuracy with a precision report of 0.49, recall of 0.48, f1 score of 0.40 for class 0 and a precision report was 0.67, recall of 0.68, f1 score was 0.67 for class 1. Then concerning gradient boost, it showed an accuracy of 67% in making predictions and the precision report was 0.61, recall 0.45, f1 score was 0.52 for 0 class and the precision report was 0.70, recall 0.82 and f1 score was 0.75 for 1 class. Then we applied XGBoost and got an accuracy of 65% with a precision report of 0.56, recall of 0.49, and f1 0.52 for class 0, and for class 1 we got a precision report of 0.70, recall of 0.75, and f1 score of 0.72. Lastly, we went for the ensemble model and achieved 68% accuracy for class 1 we got a higher score compared to class 0 and for 1 the precision report was 0.72, recall 0.78, for f1 score was 0.75 whereas for class 0 the precision report was 0.60, recall 0.52 and f1 score was 0.56.
Next, to identify the best model for predicting subscription buying behavior, the performance of these models was evaluated using a k-fold cross-validation process based on recall, accuracy, precision, f1-score, and the confusion matrix. The study shows a strong connection between social media engagement and the tendency to purchase online subscription services among youth and young adults. The machine learning models show acceptable accuracy, which proves that social media engagement through Facebook, Instagram, YouTube, etc. has a significant impact on online subscription buying behavior. This encourages the subscription providers to use social media marketing strategies to increase their subscription rates and overall customer engagement. From our surveyed data, we have seen about 50.5% who get influenced by social media recommendations buy online subscriptions one to two times monthly. Subscription services are less frequent compared to online food services as the purchasing decision depends on content quality and pricing which makes prediction less simple but we have seen about 60.7% of the respondents use different subscriptions through social media advertisements, which indicates that young users are influenced to use subscriptions because of social platforms and the impact they have from influencers' inspiration and thus we can say that our assumption on the hypothesis that social media has an impact on users decisions to buy online subscriptions is supported by the acceptable accuracy of the model. Also, our run models' predictive result approved the hypothesis that social media engagement has a powerful impact on people's decisions to purchase online subscription services.

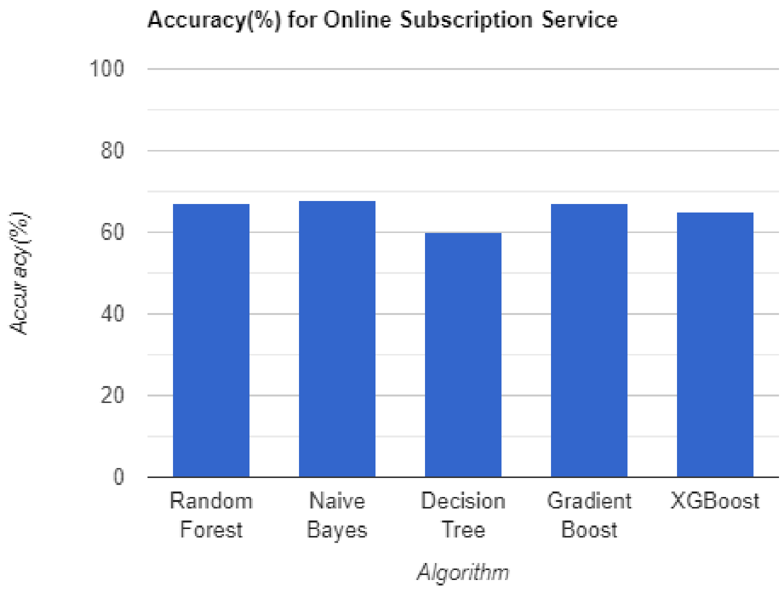

Figure 4.2: Accuracy Graph for Online Subscription Services

4.1.3 Result of Online Gaming

Table 4.3: Accuracy, Precision, Recall and F1-score of playing online games and doesn't play online games for the 6 Different ML Models

Model Name	Purchase Online Games				Doesn't Purchase Online Games			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Random Forest	0.63	0.69	0.79	0.74	0.63	0.47	0.34	0.40
Naive Bayes	0.65	0.71	0.76	0.73	0.65	0.50	0.44	0.47
Decision Tree	0.56	0.66	0.65	0.65	0.56	0.39	0.40	0.39
Gradient Boosting	0.64	0.69	0.79	0.74	0.64	0.48	0.35	0.41
XGBoost	0.62	0.69	0.75	0.72	0.62	0.45	0.37	0.41
Ensemble Model	0.63	0.70	0.76	0.73	0.63	0.47	0.40	0.43

At first, we examined the hypothesis by the data that we had collected from 1018 consumers. After that, we applied our gathered dataset to our algorithm so that we could get the desired outcome of the prediction by machine learning models. Then we have focused on some selected features that include users' frequency of usage on platforms like Facebook, Instagram, and YouTube for gaming purchase decisions, how influencers and streamers work, users' buying behavior, and the time they dedicate to game services. To examine how social media engagement has impacted online game purchasing behaviors, we have applied the algorithms, including Random Forest, Decision Trees, Gradient Boost, Naive Bayes, XGBoost, and Ensemble Model to predict the behavior of the consumer. We applied smote and k-fold for some models and after applying the Random Forest model performed by achieving 63% in its prediction of users' involvement in social media to buy online games. The f1 score was 0.40, the recall was 0.34, and the precision was 0.47 for the class of 0 and 1 precision report was 0.69, the recall was 0.79, and the f1 score was 0.74 on decision-making purposes of online game purchases. Naive Bayes produced predictions with 65% accuracy with an f1 score of 0.47, recall of 0.44, and precision report

of 0.50 for class of 0 and for class 1 precision report was 0.71, recall of 0.76, f1 score of 0.73. The Decision Trees, which provide high-level accuracy, here produced 56% accuracy with an f1 score of 0.39, recall of 0.40, and precision report of 0.39 for 0 class. Gradient Boost being a strong model here made 64% accuracy and the f1 score was 0.41, recall 0.35, and precision report 0.48 for the class of 0 and for 1 precision report was 0.69, recall 0.79, f1 score 0.74. Then we applied XGBoost and got an accuracy of 62% with a precision report of 0.45, recall of 0.37, and f1 of 0.41 for class 0, and for class 1 we got a precision report of 0.69, recall of 0.75, and f1 score 0.72. Lastly, we went for the ensemble model and achieved 63% accuracy for class 1 we got a higher score compared to class 0 and for 1 we got a precision report of 0.70, a recall of 0.76, for an f1 score of 0.73, and 0 class precision report 0.47, recall 0.40 and f1 score was 0.43. Then, to identify the best model for predicting online game service buying behavior, a k-fold cross-validation process has been used based on recall, accuracy, precision, f1-score, and the confusion matrix to test the performance of these models.
From our research, it can be stated that social media highly impacts online game purchasing decisions, specifically those who are directly connected to some social platforms. About 65.1% of gamers' respondents to the survey said that their decision to purchase games and content related to it was influenced by social media content. Also sometimes game purchasing decisions depend on the launching time of a new game, updating on a timely modified version has an effect on the accuracy of prediction. However social media engagement through Facebook, Instagram, and YouTube has an impact on online game service buying behavior and has been proven by the reliable accuracy of the machine learning models. Game developer companies get benefits from social media marketing strategies that increase their selling rates and overall customer engagement. From our surveyed data, we have seen that consumers who are influenced by social media recommendations buy online games frequently, which indicates social media users' frequency of buying online services. Also, the used models' predictive outcome supports the hypothesis that social media engagement has a powerful impact on youths' decisions to purchase online game services.

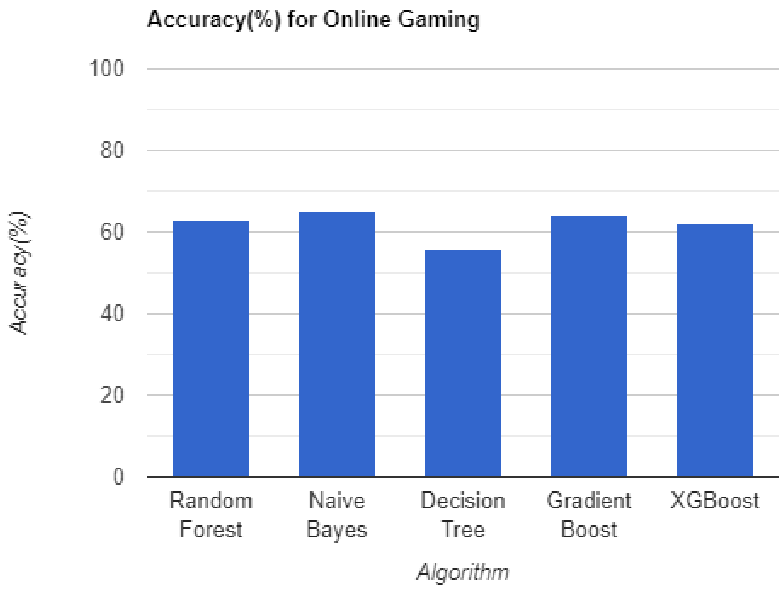

Figure 4.3: Accuracy Graph for Online Game Purchase

Chapter 5

Conclusion and Future Work

5.1 Limitations and Future Research Directions

There are some limitations which are pertinent to the project. Our analysis has shown that it can detect if a buyer is influenced by social media or not. The results that we achieved depend on the data, which was derived from the experiences of different consumers. However, we were unable to collect data from the sellers, and we were not able to focus on the results from a business or seller perspective. Furthermore, we conducted this research by focusing only on consumer experience. Thus, if in the near future, we witness a new revolution in the social media sphere, then the collected dataset will not be able to properly detect/predict. Also, we would then have to recollect data based on the new change that has affected the social media sphere. Moreover, the focus of our research was youths' and young adults' buying preferences, but there is a large number of consumers who are older than 30 and use social media. In the future, we will implement data from people who are aged above 30 in our research and try to predict their preferences. In addition, the seller's behavior will also be implemented in the dataset to understand their marketing strategy to influence consumers. Our methods and algorithms don't consider human emotions or sentiments. As our research progressed step by step, we realized that the sentiment of people also has somewhat of an impact on their decision-making and affects their product purchase decisions. Here we have thought about the implementation of the BART algorithm, which is an NLU (natural language understanding) algorithm used to detect if the consumer is influenced by social media or not by utilizing their social media posts as the dataset. Another major limitation is the accuracy level of the subscription purchase and online gaming target variables are lower than our expectation. To improve we implemented a new method like RFE but could not get any better results. Primarily we think that the reason for this low accuracy is irrelevant data provided by respondents. Thus, in the future, we would like to improve this aspect of our study.

5.2 Conclusion

In conclusion, our study focused on consumer behavior in the digital age with an emphasis on how social media is incorporated into daily life. We gained a crucial understanding of consumer preferences and interactions by utilizing machine learning. It came out that factors like entertainment, customization, and word of mouth greatly affect purchasing decisions. Additionally, demographic details like age are very important in influencing consumer choices, particularly in online shopping. Our study shows the potential for data analysis and predictive modeling to improve marketing strategy. Businesses can improve customer experiences in the dynamic world of social media by adapting methods based on these factors. The importance of strategies based on data for companies aiming to succeed in the modern consumer market is highlighted by this study.

Bibliography

[1] C. A. Jothi and A. M. Gaffoor, "Impact of social media in online shopping," *Journal on Management Studies*, vol. 3, no. 3, pp. 576–586, 2017.

[2] S. Satpathy and S. Patnaik, "Role of social media marketing on consumer purchase behaviour: A critical analysis," in *2019 International Conference on Applied Machine Learning (ICAML)*, IEEE, 2019, pp. 92–97.

[3] B. S. Arasu, B. J. B. Seelan, and N. Thamaraiselvan, "A machine learning-based approach to enhancing social media marketing," *Computers & Electrical Engineering*, vol. 86, p. 106 723, 2020.

[4] B. C. Weilun Wang Goutam Chakraborty, "Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm," 2020.

[5] K. Chaudhary, M. Alam, M. S. Al-Rakhami, and A. Gumaei, "Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics," *Journal of Big Data*, vol. 8, no. 1, pp. 1–20, 2021.

[6] G. T. Ioannis Mollas Nick Bassiliades, "Conclusive local interpretation rules for random forests," 2021.

[7] B. Madani and H. Alshraideh, "Predicting consumer purchasing decision in the online food delivery industry," *arXiv preprint arXiv:2110.00502*, 2021.

[8] P. Ebrahimi, M. Basirat, A. Yousefi, M. Nekmahmud, A. Gholampour, and M. Fekete-Farkas, "Social networks marketing and consumer purchase behavior: The combination of sem and unsupervised machine learning approaches," *Big Data and Cognitive Computing*, vol. 6, no. 2, p. 35, 2022.

[9] V. Parihar and S. Yadav, "Comparative analysis of different machine learning algorithms to predict online shoppers' behaviour," *International Journal of Advanced Networking and Applications*, vol. 13, no. 6, pp. 5169–5182, 2022.

[10] Z. Xu, G. Zhu, N. Metawa, and Q. Zhou, "Machine learning based customer meta-combination brand equity analysis for marketing behavior evaluation," *Information Processing & Management*, vol. 59, no. 1, p. 102 800, 2022.

[11] N. Montalvo, "When representation learning meets random forest: Deep neural decision forest.," 2023.

[12] G. Boesch, "Ensemble learning: A combined prediction model (2025 guide)," 2024.

[13] S. Saridena, "Naive bayes," 2024.

[14] Sylen, "Machine learning - classification model gradient boosting machine," 2024.

[15] M. A. H. Rafath and A. Sultana, "Consumer behavior analysis of social media networks by using machine learning,"