Behavioural Analysis of Factors Influencing Online Food Ordering and Its Relation to Obesity

S. M. Fazle Rabby Labib[a,*], Fahmida Zaman Achol[a], Md. Aqib Jawwad[a], Dr. Md. Golam Rabiul Alam[a] and Dr. Ashis Talukder[b]

[a] *BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh*
[b] *University of Dhaka, Nilkhet Rd, Dhaka 1000, Bangladesh*

ARTICLE INFO

Keywords:
Online Food Ordering
Obesity
Statistical Analysis
Prediction Models
Customer Segmentation
Hypothesis Testing

ABSTRACT

The purpose of this research is to determine what influences people to order food online and the intensity of ordering as well as to see if there exists any correlation between online food ordering and obesity among the people of Bangladesh. We conduct an online questionnaire survey analyzing data from 343 participants aged 16-30 residing in Bangladesh. We use Ordinary Least Squares, Decision Tree, and Random Forest to determine the factors influencing customers' decision to order food online. The most influential factors are impulsive decisions, convenience of ordering, variety of options available, and promotional offers and discounts. Then, we conduct statistical analyses to determine the correlation between obesity and the intensity of online food ordering. We find a weak positive relationship (p-value = 2.43^{-7}, coefficient = 0.28). We classify frequent users of OFO by using classifiers such as Logistic Regression, Naive Bayes, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Machine and K-Nearest Neighbors, where we find Random Forest to perform the best with an accuracy of 81%. Customer segmentation using K-Means clustering reveals that infrequent orders perceive OFO services as expensive, while frequent customers find the food lacking in nutrition. Several Chi-Squared Tests are conducted, revealing significant associations. Firstly, tech savviness is strongly related to the perception of OFO services as user-friendly. Secondly, individuals leading busy lives tend to make impulsive decisions. Lastly, the affordability of OFO services correlates with ordering behavior driven by promotional offers and discounts. These insights, along with the developed prediction models and customer segmentation, contribute to a deeper understanding of consumer behaviour in an increasingly digitalized food landscape.

1. Introduction

Online Food Ordering (OFO) Services have transformed our access to meals, offering convenience unparalleled by traditional dining. However, this convenience comes with health concerns, notably the rise in major health issues like high blood pressure, obesity, diabetes, cardiovascular diseases, and various types of cancer. Obesity, in particular, has become a critical global health challenge. It is attributed to an abnormal accumulation of body fat that significantly impacts overall health and well-being. The WHO [1] reports a tripling of obesity rates since 1975, marking it a pressing public health problem worldwide.

The surge in OFO services has particularly amplified during COVID-19, redefining how we procure meals. With market values skyrocketing globally, including within Bangladesh, platforms like foodpanda, HungryNaki, and Pathao Food have become integral parts of modern food consumption. Yet, what influences individuals to use these services and how this behaviour might relate to the rise in obesity remains largely unexplored.

The linkage between the popularity of OFO services and the escalating rates of obesity is a pressing concern. According to the National Health Service (NHS) [2] obesity stems from excessive calorie intake and insufficient

*Corresponding author. Phone: +880 1623925656. Department of Computer Science & Engineering, Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh

s.m.fazle.rabby.labib@g.bracu.ac.bd (S.M.F.R. Labib); fahmida.zaman.achol@g.bracu.ac.bd (F.Z. Achol); md.aqib.jawwad@g.bracu.ac.bd (Md.A. Jawwad); rabiul.alam@bracu.ac.bd (Md.G.R. Alam); ashis@du.ac.bd (A. Talukder)

www.bracu.ac.bd (S.M.F.R. Labib); www.bracu.ac.bd (F.Z. Achol); www.bracu.ac.bd (Md.A. Jawwad); www.bracu.ac.bd (Md.G.R. Alam); www.du.ac.bd (A. Talukder)

ORCID(s): 0009-0008-5397-3881 (S.M.F.R. Labib); 0009-0001-0043-5798 (Md.A. Jawwad); 0000-0002-9054-7557 (Md.G.R. Alam); 0000-0003-2991-9136 (A. Talukder)

physical activity. Most OFO platforms offer highly processed, calorie-rich food options, lacking vital nutritional information, potentially contributing to this concerning trend.

Our research aims to bridge this gap by investigating the factors influencing OFO service usage and its potential correlation with rising obesity rates, particularly among the Bangladeshi populace. To do that we need a significant amount of data. A problem arises as there is little to no data available related to this problem worldwide let alone in Bangladesh. As a result, the research is trying to address the following questions:

What factors influence users' decisions to order food through online services, and does the intensity of ordering correlate with the increase in obesity rates among the people of Bangladesh? Is there a relationship between these factors? Additionally, can users be effectively classified and segmented into multiple clusters based on their behaviour?

In this research, we create a novel dataset on Bangladeshi customers who use OFO services to investigate the factors that influence the use of OFO services. We examine potential correlations between Body Mass Index (BMI) and monthly order frequencies. Moreover, we develop a predictive model that classifies frequent online food orders. Employing customer segmentation techniques, we bifurcate the user base into two clusters for in-depth analysis. Finally, we conduct hypothesis testing to find the relation between the factors. Importantly, such research is unprecedented within the Bangladeshi context. Our primary contributions can be summarized as follows:

1. We identify the factors driving the adoption of OFO services among Bangladeshi consumers, utilizing techniques such as Decision Tree, Random Forest, and Ordinary Least Squares to assess feature importance.
2. We conduct hypothesis testing to determine the relation among influencing factors using Chi-squared test of independence.
3. We have found that even though it is very weak, there exists a correlation between obesity and orders per month. We also conducted a correlation analysis between other BMI categories and orders per month as well as between physical activity, order history and orders per month. For this, we used Pearson Correlation, Spearman's Rank Correlation and Point Bi-serial Correlation.
   We explore the correlation, albeit weak, between BMI and monthly order frequencies using various correlation analyses such as Pearson, Spearman's Rank, and Point Bi-serial correlations. Additionally, we examine the correlations among BMI categories, physical activity, order history, and order frequencies.
4. We develop a model to classify frequent users of OFO services, employing classifiers such as Logistic Regression, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting Machine, and K-Nearest Neighbors.
5. We apply customer segmentation techniques, specifically K-Means and PCA, to find user clusters, offering insights into consumer behaviour.
6. We create a novel dataset of Bangladeshi OFO service users, a significant addition to the Bangladeshi research landscape.

2. Literature Review

The surge in online food ordering (OFO) has reshaped consumer habits, which brings us to explore its multifaceted impact. Here, we examine various studies related to OFO, encompassing its influence on consumer behavior, health implications like obesity, and the factors driving its popularity.

Kale *et al.* [3] highlight convenience and flexibility as prime reasons behind the preference for using OFO which is supported by Alagoz *et al.* [4] who emphasize the significance of an easy ordering process, particularly among university students. Parameshwaran *et al.* [5] and Anam *et al.* [6] delve into the influence of advertising, social factors and the shift to online platforms, accentuated during the COVID-19 pandemic.

In examining health ramifications, Chatterjee *et al.* [7] use machine learning methods to explore obesity risk factors. They emphasize the need for understanding lifestyle choices of consumers. Prenctice *et al.* [8] shed light on the role of energy-dense food in obesity, stressing its prevalence in fast food offerings and its adverse impact, especially among children. Meanwhile, studies by Harahap *et al.* [9] and Kurniawati *et al.* [10] emphasize the correlation between online food ordering, dietary choices, and obesity risks among university students, offering insights into nutritional aspects and consumption patterns, albeit with differing conclusions. Whereas, Harahap *et al.*suggests OFO and fast food consumption leads to over consumption which contributes to obesity, Kurniawati's study reveals the nutritional content of the meals ordered online does not establish a direct link between OFO and nutritional state of college students.

Studies conducted by Hong *et al.* [11] and Ayubi *et al.* [12] analyze customer intentions, behavioral aspects and influences across demographics and regions. Similarly, Jahidi *et al.* [13] and Inthong *et al.* [14] concur, highlighting the potential variation in motivations across different regions.

Despite extensive research, gaps persist, particularly regarding the Bangladeshi perspective. Existing studies primarily hail from diverse geographical locations, with datasets unavailable for future research. This gap underscores the need for localized studies in Bangladesh to understand cultural nuances and their influence on OFO dynamics.

In summary, while existing studies provide invaluable insights into OFO's global impact, there's a clear need for localized research in Bangladesh. Understanding how OFO affects Bangladeshi consumers, considering cultural, socioeconomic, and dietary differences, will be crucial for informing policies and practices in this evolving landscape.

3. Methodology

In this section, we discuss the detailed methodology of our research work.

In figure 1, we can see the top level overview of our proposed system. We start by collecting responses from the survey and constructing our dataset. We do data pre-processing and test for validity and reliability. Then we do correlation analysis, find factors that influence the use of OFO services, create a prediction model, apply customer segmentation on our data and do some hypothesis testing. Finally, we show our findings.

3.1. Data Collection

In order to investigate the factors that influence the decision to order food online among individuals in Bangladesh, a survey is conducted. The survey questions are divided into two parts. One is the demographic data and the other is questions related to online food ordering and the factors influencing it. There were in total 25 questions in the questionnaire. The survey is done anonymously to protect the respondent's privacy. The survey is done through Google Forms and is designed to take the smallest amount of time possible. To accomplish that, we first share the form with our peers and take feedback from them. We incorporate the feedback and make it public. The questions ranged from text-based, multiple-choice questions (MCQ), check boxes, and Likert scale questions. The survey questions are shared in the appendix.

Throughout the data collection period we receive several queries regarding why a respondent needs to login to their email to respond to our questions. We explain it to them that this is necessary to protect the survey from getting spam and random responses. Another concern we faced, is about the user being reluctant to share their personal information like height and weight we assured them that none of the answers are connected to their emails. We also made sure to not force anyone to share their information and only took responses if they gave it willingly.

3.2. Data pre-processing

After collecting the survey responses, the column names are modified for improved readability and ease of use. The height variable, which were initially recorded in feet and inches, get converted to meters for consistency. Using the converted height and weight variables, the BMI of each respondent is calculated. A new variable, BMI_CAT, is created to store the categorical values of BMI, including "Underweight", "Normal", "Overweight" and "Obese". In order to maintain the focus on the target population of individuals aged 16-30, any responses from individuals outside of this age range are removed, as about 98% of the data consists of individuals within this age group. The Likert scale data obtained from the questions related to various factors are encoded into ordinal numerical values for ease of analysis. This conversion allowed for a more accurate statistical analysis of the collected data. We remove any responses where orders per month is less than 1, since those information are not useful for our research problem. After pre-processing we end up with valid 331 responses which we use to conduct rest of our research.

3.3. Dataset Validity and Reliability

Before applying all of the discussed methods we use Chronbach's Alpha, which is a way of assessing reliability by comparing the amount of shared variance, or covariance, among the items making up an instrument to the amount of overall variance. This is done to test the validity and reliability of our dataset. The idea is that if the instrument is reliable, there should be a great deal of covariance among the items relative to the variance.

After applying Chronbach's Alpha on our dataset we get the value of 0.89. According to this paper by Taber [15] a chronbach's alpha value higher than 0.70 is considered acceptable and above 0.80 and above is considered better. Therefore we can safely say that our dataset is valid and reliable.

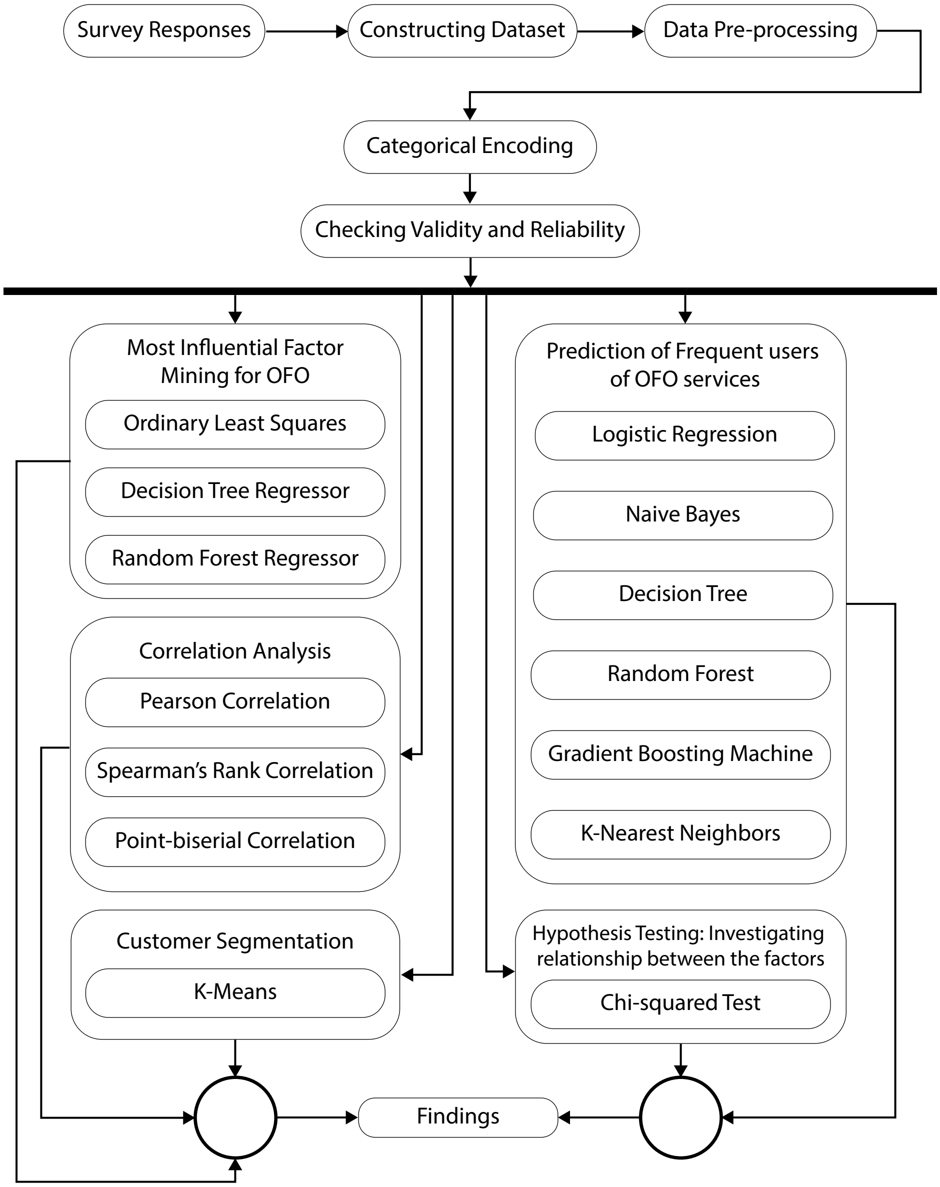

Figure 1: Top level overview of the proposed system.

3.4. Most Influential Factor Mining for OFO

To find factors that are influencing a person to use OFO, we use Ordinary Least Squares (OLS) [16], Decision Tree Regressor [17] and Random Forest Regressor [18].

Ordinary Least Squares: This is a linear least squares method that chooses unknown parameters in a linear regression. This method tries to find the line that fits the data points the best by reducing the total amount of error between the predicted values and the actual values of the dependent variable. The model takes the form of an equation:

$$y = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n \tag{1}$$

where y is the dependent variable, x_1, x_2, ..., x_n are the independent variables, and b_0, b_1, b_2, ..., b_n are the coefficients of the equation. The goal of OLS is to find the values of the coefficients that minimize the difference

Table 1
Questionnaire Description.

Attributes	Data Type
Gender	Categorical Value
Age	Numeric Ranges
Weight	Numeric Value (Kilogram)
Height	Numeric Value (Feet and Inches)
Educational qualification	Categorical Value
Employment status	Categorical Value
Financial dependency	Categorical Value
Marital status	Categorical Value
Physical activity	Numeric Range: 1-7 (per week)
Preferred OFO apps	Categorical Value
OFO apps use duration	Categorical Value
Orders per month	Numeric Value
Ordering time	Categorical Value
Types of food ordered	Categorical Value
Tries out new technologies	Likert Scale
Makes impulsive decisions	Likert Scale
Too busy to cook	Likert Scale
Does not like to cook	Likert Scale
Ordering online is easy	Likert Scale
Variety of options	Likert Scale
Ordering is inexpensive	Likert Scale
Promo codes and discounts	Likert Scale
Food is delivered quickly	Likert Scale
Food is safe and hygienic	Likert Scale
Food is nutritious	Likert Scale

between the predicted values of y (based on the equation) and the actual values of y. A system of linear equations are solved to find the coefficients. The matrix representation of this system is $X'Xb = X'y$, here X is the matrix of independent variables, b is the vector of coefficients and y is the vector of dependent variable. These coefficients can tell us how strong and which way the independent variable affects the dependent variable. The one with the biggest coefficient is the most important factor in influencing the dependent variable.

Decision Tree: This method uses a tree-shaped model to make decisions and predict outcomes. The tree is created by splitting the data into smaller groups based on the values of the independent variables. This process is repeated recursively until we have subsets that can predict the consequences of each decision. At each internal node of the tree, a decision rule is applied to the data, and the tree branches accordingly. Once the tree is built, we used it to identify the most important variables for ordering food online by looking at the variables that are used in the decision rules at the top of the tree.

Random Forest: This is an ensemble method that merges the predictions of multiple decision trees. To use this method, we first divide the data into different subsets. Then, we train multiple decision trees on each subset. Each tree makes its own predictions based on the data it was trained on. Finally, we take the average of all the predictions from these trees to get the final result. Random forest is more robust to overfitting than a single decision tree, and it also allowed us to identify the most important variables by looking at the variables that are used in the decision rules of the majority of the trees.

The most important equation for Random Forest is the feature importance formula which is used to calculate the importance of each feature. The feature importance of a feature is calculated as the average decrease in impurity across all decision trees in the forest. The impurity can be calculated using the Gini Impurity.

$$Gini = 1 - \sum_{i=1}^{m} p_i^2 \tag{2}$$

We use the 11 factors which are liking to try new technologies, making impulsive decision, being too busy too cook, disliking towards cooking, finding ordering online to be easy, having variety of options, finding ordering to be inexpensive, because of promotional offers and discounts, quick delivery, food safety and nutritious value. We use orders per month as the dependent variable. We calculate mean and standard deviation of each factors. Then apply OLS, Decision Tree Regressor and Random Forest Regressor. After that, we use the r-squared value to measure the goodness of fit of our regression models. The r-squared value for OLS, Decision Tree and Random Forest are 0.64, 0.01 and 0.04. Which means OLS is a better approach for our data.

3.5. Hypothesis Testing: Investigating relationship between the factors

In our study, we also conduct some hypothesis testing using the chi-squared test [19].

Chi-Squared Test: This test is used to determine if there is a significant difference between an observed distribution and a theoretical distribution. The test statistic is calculated using the following equation:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

Where:
O = observed frequency
E = expected frequency

The degrees of freedom for the test is calculated as:

$$df = (number\ of\ rows - 1) * (number\ of\ columns - 1) \tag{4}$$

The p-value is then looked up in a chi-squared distribution table with the corresponding degrees of freedom. If the calculated p-value is less than the chosen significance level (usually 0.05), then the null hypothesis (that there is no significant difference between the observed and theoretical distributions) is rejected, and it is concluded that there is a significant difference between the two. We use the chi-squared test of independence as the factors have ordinal values. So, something like Pearson's correlation coefficient might not be the best measure of testing here as it assumes that the data is normally distributed and continuous.

We test the following hypotheses for our study,

People who are open to experimenting with new technology such as OFO services, may find it to be effortless to place an order. These platforms offer easy to use and intuitive interface which simplifies the ordering process, thus allowing users to place orders quickly without any sort of hassle. Moreover, these applications often provide the option to track orders in real time, personalized recommendations. All of these features can enhance the overall ordering experience and make it more convenient for the users to use. Based on this argument we come up with the following hypothesis:

H1: There is a significant relationship between trying out new technology and thinking that using OFO service is easy.

People who have busy schedules may be more likely to make impulsive decisions when it comes to using OFO services. Due to their busy lifestyle, they may not have the energy to shop for ingredients or cook at home. Therefore, they may turn to quick and easy options such as ordering takeout. Based on this argument we come up with the following hypothesis:

H2: There is a significant relationship between making impulsive decisions when using OFO service and being too busy to cook.

Table 2
Results of Hypothesis Testing.

Serial	P-Value	Decision
H1	< 0.001	Accepted
H2	0.114	Rejected
H3	< 0.001	Accepted
H4	< 0.001	Accepted

People who do not like to cook may appreciate the variety of options that are available on the OFO platforms. For them, the thought of preparing meals at home can be unappealing or they might prefer to have a variety of options to choose from when ordering food. Based on this argument we come up with the following hypothesis:

H3: There is a significant relationship between people not liking to cook and finding there are many options to choose from when using OFO service.

Promotional offers or discounts can make the use of OFO services more cost-effective for users. These offers can take various forms such as coupons, loyalty programs, and special deals. These can be used to reduce the cost of items, delivery fees or sometimes even the entire order. By making use of these discounts, customers can save money on their orders. Therefore leading them to believe that ordering food online is inexpensive. Based on this argument we come up with the following hypothesis:

H4: There is a significant relationship between thinking OFO service is inexpensive and ordering food because of promo codes and discounts.

3.6. Correlation Analysis

To examine the relationship between BMI (x) and the frequency of ordering food online (y), we apply several correlation analysis methods. These are Pearson Correlation [20] where we use the actual BMI value of a respondent, Spearman's Rank Correlation [21] where the BMI categories are used and Point-biserial Correlation [22] where we convert BMI categories to binary categories.

Pearson correlation: This method measures the linear relationship between two continuous variables. The process involves calculating the means ($\bar{x}, \bar{y}$) and standard deviations of each variable, followed by computing the product of the deviations of each variable from its mean for each data point. The following equation is then used to calculate Pearson's correlation coefficient, r:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

The resulting value of r ranges from -1 to 1, here -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship. To evaluate the significance of the correlation coefficient, a p-value is determined using the correlation coefficient, degrees of freedom (df = n-2) and a significance level of $\alpha = 0.05$ is used usually. The table of critical values is used to find the p-value. The correlation coefficient is assumed as zero by the null hypothesis, implying that there is no correlation between the two variables. The alternative hypothesis is that the correlation coefficient is not zero, implying that there is a correlation between the two variables. If the p-value is less than α, we get enough evidence to accept the alternative hypothesis and reject the null hypothesis. This method is applicable when both variables are continuous and assumes the data follow a normal distribution. It also assumes the observations are independent of each other.

Spearman's rank correlation coefficient: This is a non-parametric way of measuring correlation. With this we can measure how well the relationship between two variables can be described by using a monotonic function. It is similar to Pearson correlation but instead of measuring linear association, this evaluates rank order similarity between variables.

Table 3
Pearson and Spearman Correlation Test Results of BMI vs Orders Per Month.

Test	P-value	Coefficient	Relationship
Pearson	2.43^{-7}	0.28	Weak Positive
Spearman	0.003	0.15	Very Weak Positive

In this method each variable is assigned ranks based on their values. The smallest value receives a rank of 1 and the second smallest a rank of 2 and so on. Then it determines difference between ranks for each paired observation and these differences represent the change in the variables' relative positions. Each difference obtained in previous step is then squared to stop the positive and negative differences canceling each other out. Finally, the coefficient is calculated using the following formula:

$$\rho = 1 - \frac{(6 * sum\ of\ squared\ differences)}{n * n^{2-1}} \tag{6}$$

The results will range from -1 to 1, where a value of +1 indicates that as one variable increases, the other variable also increases consistently. On the other hand, a value of -1 means that as one variable increases, the other variable decreases consistently. A coefficient of 0 suggests no monotonic relationship between the variables. This approach only works when the variables are ordinal or ranked. This method is also less sensitive to outlier data.

Point-biserial correlation: This approach measures the strength and direction of association between a continuous variable and a binary variable.

At first, the data is divided into two groups based on the binary variable. One group represents the presence of the characteristic while the other group represents the absence of it. Then, the mean value of the continuous variable is calculated for each group separately. After that, standard deviation of the whole dataset is calculated, disregarding the grouping which represents the variability of the continuous variables across the entire dataset. The mean of the group representing the absence of the characteristic is then subtracted from the mean of the group representing the presence of it. We divide the difference by the standard deviation. We then get the point-biserial correlation coefficient which quantifies the strength and direction of the relationship between the continuous and binary variable.

The coefficient value ranges from -1 to +1. A positive value indicates a positive relationship, meaning higher values of the continuous variable tend to be associated with the presence of the characteristic. A negative value indicates a negative relationship, where higher values of the continuous variable tend to be associated with the absence of the characteristic. A value of 0 indicates no relationship between the continuous variable and the binary variable. This approach assumes that the continuous variable follow a normal distribution or at least approximate normality. Skewed or variable with outliers may affect the accuracy. The observations should also be independent of each other. Larger sample sizes tend to yield more accurate estimates.

For Pearson correlation we use the BMI with continuous values and orders per month which also consists of continuous values. When using Spearman's rank correlation coefficient we use the categorical values of BMI instead and for Point-biserial correlation, we divide the BMI categories in binary groups such as not obese and obese, not normal weight and normal weight and apply the method.

We also apply Pearson and Spearman's correlation on physical activity and orders per month. Treating the values for physical activity as continuous for Pearson and categorical for Spearman's.

3.7. Prediction of Frequent users of OFO services

To create classify frequent users of OFO services we use several methods such as Logistic Regression [23], Naive Bayes [24], Decision Tree Classifier [17], Random Forest Classifier [18], Gradient Boosting Machines (GBM) [25] and K-Nearest Neighbour (KNN) [26].

Logistic Regression: This is a commonly used statistical model for tasks that involve classifying things into two categories. The main aim is to predict the probability of an instance belonging to a specific class. Unlike linear regression, which predicts continuous values, it is specifically designed to predict the probability of an event happening. Instead of providing a direct numerical output, this method calculates the likelihood of an event occurring based on the input variables. This makes it suitable for tasks where we are interested in determining the likelihood or probability of a specific outcome.

Logistic Regression models the relationship between the predictor variables and the output class probabilities using a logistic function or sigmoid function. The sigmoid function is defined as:

$$S(z) = \frac{1}{1 + e^{-z}} \tag{7}$$

Here, z represents a linear combination of the input features and model parameters.

The model is trained by finding the optimal values for the model parameters which minimize the error between the predicted probabilities and the actual class labels in the training data. This is typically done using optimization algorithms such as Newton's method or Gradient Descent. The objective is to minimize the log loss of the predicted probabilities or maximize the likelihood of it. Once the model is trained and the optimal parameters are obtained, we can use the model to make predictions on new, unseen data. The sigmoid function is applied to the linear combination of the input features and parameters to obtain the predicted probability. A common threshold (0.5) is then used to classify the instance into one of the two classes based on the predicted probability. If the predicted probability is above the threshold, it is classified as class 1; otherwise, it is classified as class 0.

Naive Bayes: This algorithm is built upon the Bayes' theorem and assumes that given the class label, the features are conditionally independent of each other.

This algorithm starts by calculating the prior probabilities of each class in the dataset. The prior probability of a class is the probability of encountering that class in the dataset without considering any feature information. For each feature in the dataset, this method calculates the conditional probability of observing that feature given a specific class label. This is done by counting the occurrences of each feature in the training samples belonging to a particular class. Once the prior probabilities and conditional probabilities are computed, Bayes' theorem is applied to calculate the posterior probability of each class given the observed features. Finally, the classifier predicts the class label for a new, unseen sample by selecting the class with the highest posterior probability.

The algorithm for Decision Tree and Random Forest classifier works in the same as the regressor approach but the classifier is more suitable for categorical or discrete class labels.

Gradient Boosting Machines: This algorithm combines multiple weak predictive models (usually decision trees) to create a better predictive model. It is an ensemble learning method that sequentially builds an additive model by minimizing a predefined loss function using gradient descent.

A weak learner which usually is a decision tree with a small depth which is also called a decision stump, is fitted to the training data. The tree is trained to minimize the loss function with respect to the target variable. The predictions of the weak learner are then subtracted from the true values of the target variable to obtain the residuals. The residuals represent the errors or the parts of the target variable that the model has not yet captured. Another weak learner is then fitted to the residuals obtained from the previous step. The goal is to find a new model that can capture the remaining patterns in the data that the previous model missed. The new weak learner is now added to the ensemble by combining it with the previous models. To combine the models, a weight is assigned to each weak learner, this is also known as the learning rate. The weight determines the contribution of each model to the final prediction. The process of calculating residuals, fitting the next learner and updating of the model is repeated iteratively for a specified number of times or until a predefined stopping criterion is met. In each iteration, a new weak learner is fitted to the negative gradients of the ensemble model built so far. The final prediction is gained by summing the predictions of all the weak learners with their respective weights. The model assigns higher weights to the more accurate learners, which allows it to give more importance to their predictions.

K-Nearest Neighbour: KNN is an algorithm that does not rely on any assumptions about the underlying data distribution. The fundamental idea behind KNN is to classify or predict the target variable of a new data point based on the majority vote or the average of the target variables of its K number of nearest neighbors in the feature space.

At first, the value of K is figured out, which represents the number of nearest neighbors to consider for classification or regression. The optimal value of K can be found using techniques such as cross-validation or grid search. To classify or predict a new data point, the distance between that point and every other data points in the training data is calculated. The distance can be calculated using different metrics such as Manhattan distance and Euclidean distance. The distances

Table 4
Results of feature selection.

Chi-squared Test	ANOVA F-value
Gender	Gender
BMI	BMI
Employment Status	Employment Status
Financial Dependency	Financial Dependency
Physical Activity	Marital Status
Ordering History	Physical Activity
Impulsive Decision	Ordering History
Busy To Cook	Impulsive Decision
Dont Like To Cook	Busy To Cook
Ordering Easy	Dont Like To Cook
Options To Choose	Ordering Easy
Ordering Inexpensive	Options To Choose
Promo Discounts	Promo Discounts
Quick Delivery	Quick Delivery
Food Safety	Food Safety
Nutritious	Nutritious

are then sorted in ascending order and the K data points are selected with the shortest distances as the nearest neighbors. Then, the class label is determined that occurs most frequently among the K nearest neighbors. This can be done by counting the occurrences of each class label and selecting the one with the highest count. Once the class label or predicted value has been determined, it is assigned as the output for the new data point.

We divide the orders per month variable into two category. One with less or equal to 5 orders per month which covers 56% of the sample and another with more than 5 orders per month which covers the rest of 44% of the sample. We then use this as the target label against all the features available in our dataset. We scale the data, select the best features using correlation heat map and implement the algorithms. We use uni variate methods such as chi-squared test and ANOVA F-value to do some feature selection.

From table 4, we can see that 15 out of the 16 selected features are common for both methods. Therefore, we use select these 15 to build our prediction models.

We use several evaluation metrics like accuracy score, precision, recall and F1-score to find the model with best results.

3.8. Customer Segmentation

Customer segmentation is a technique in data analysis to group customers based on their similarities and differences. One way of doing this is using K-Means clustering algorithm [27], which is an unsupervised machine learning algorithm that aims to divide a given dataset into K distinct clusters, where each data point goes to the cluster with the nearest mean value. Principal Component Analysis (PCA) [28] is a technique that can be used to reduce the dimension of the dataset.

K-Means: The algorithm works by assigning data points iteratively to the cluster with the closest centroid and then updating the centroids to be the average of the data points in each cluster. This process continues until the centroids no longer change significantly.

Number of clusters K is usually chosen by using elbow method or silhouette scores. Then the the centroids of the clusters are initialized. This is done randomly or by using some other method, such as k-means++. Each data point then gets assigned to the cluster with the closest centroid. The centroids get updated to be the average of the data points in each cluster. This is repeated until the centroids no longer change significantly.

Principal Component Analysis: This is a widely used statistical technique for dimensionality reduction and data analysis. It aims to find the directions or principal components in which the data varies the most and represents the data in a lower-dimensional space while preserving the essential information.

At first, dataset is standardized to have a unit variance and zero mean. This step ensures that all features are on a similar scale and prevents variables with large variances from dominating the analysis. Then the covariance matrix of the standardized data is computed. The covariance matrix indicates the relationships between different features and measures how they vary together. The value in the i^{th} row and j^{th} column of the covariance matrix represents the covariance between the i^{th} and j^{th} features. An eigendecomposition is performed on the covariance matrix to find its eigenvalues and corresponding eigenvectors. The amount of variance explained by each principal component is represented by the eigenvalues. The eigenvectors represent the directions of these components. The eigenvalues indicate the relative importance of each principal component. PCA typically sorts the eigenvalues in descending order and selects the top-k eigenvectors corresponding to the largest eigenvalues, where k is the desired number of dimensions in the reduced space. The selected eigenvectors form a new coordinate system. The original data is projected onto this new coordinate system to obtain the reduced-dimensional representation. This projection involves multiplying the standardized data matrix by the eigenvector matrix, resulting in a transformed dataset with reduced dimensions.

For this, we use only the 11 factors we developed that influence OFO to do customer segmentation. We use K-Means [27] algorithm to perform this. We apply this using both Principal Component Analysis (PCA) [28] and one without using it. Before applying PCA, we use Kaiser-Meyer-Olkin test (KMO) [29] and Bartlett's Test of Sphericity [30] to check if the data is suitable to apply PCA. The results of KMO test gives us 0.85 which makes it sufficient to perform factor analysis according to this paper [31]. The result of Bartlett's Test of Sphericity gives us a p-value of 1.06^{-228} which is lower than the α value of 0.05 and a high chi-squared value of 1301.43. Therefore, we can apply PCA on our data.

We use elbow method and silhouette score to find optimal number of clusters. From figures 2, 4, 3 and 5 we can see that the optimal number of clusters is 2.

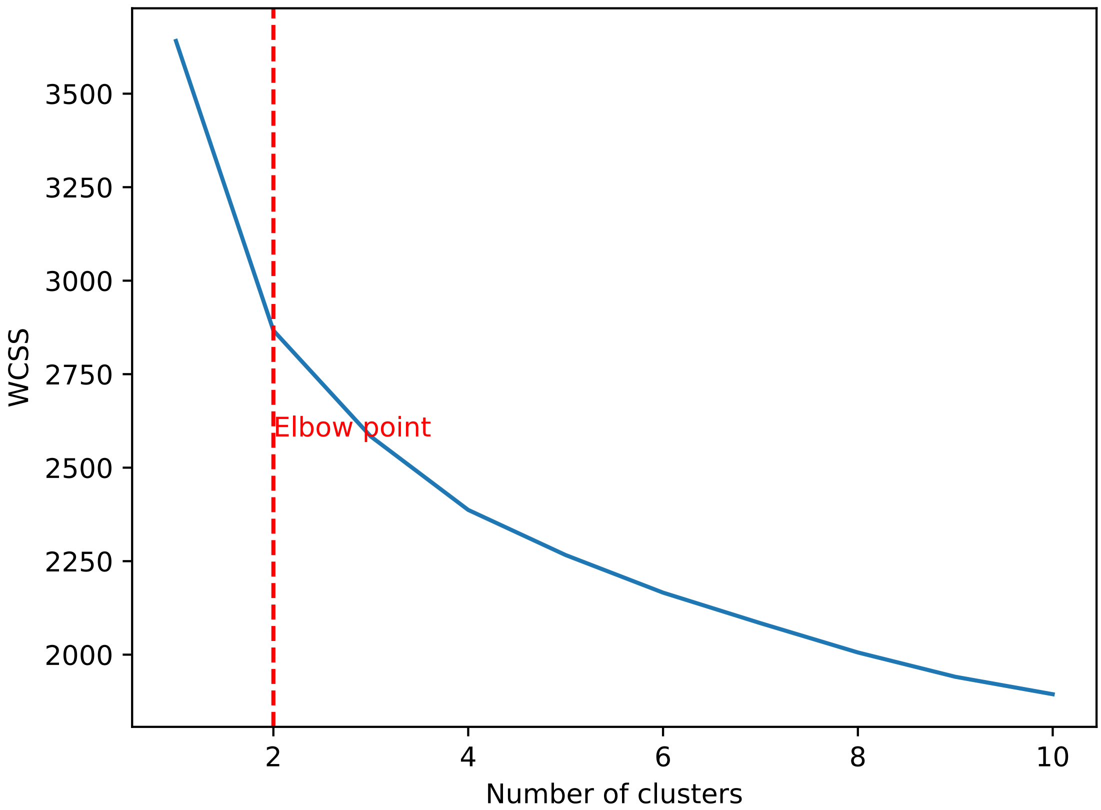

Figure 2: Elbow method for determining the optimal number of clusters.

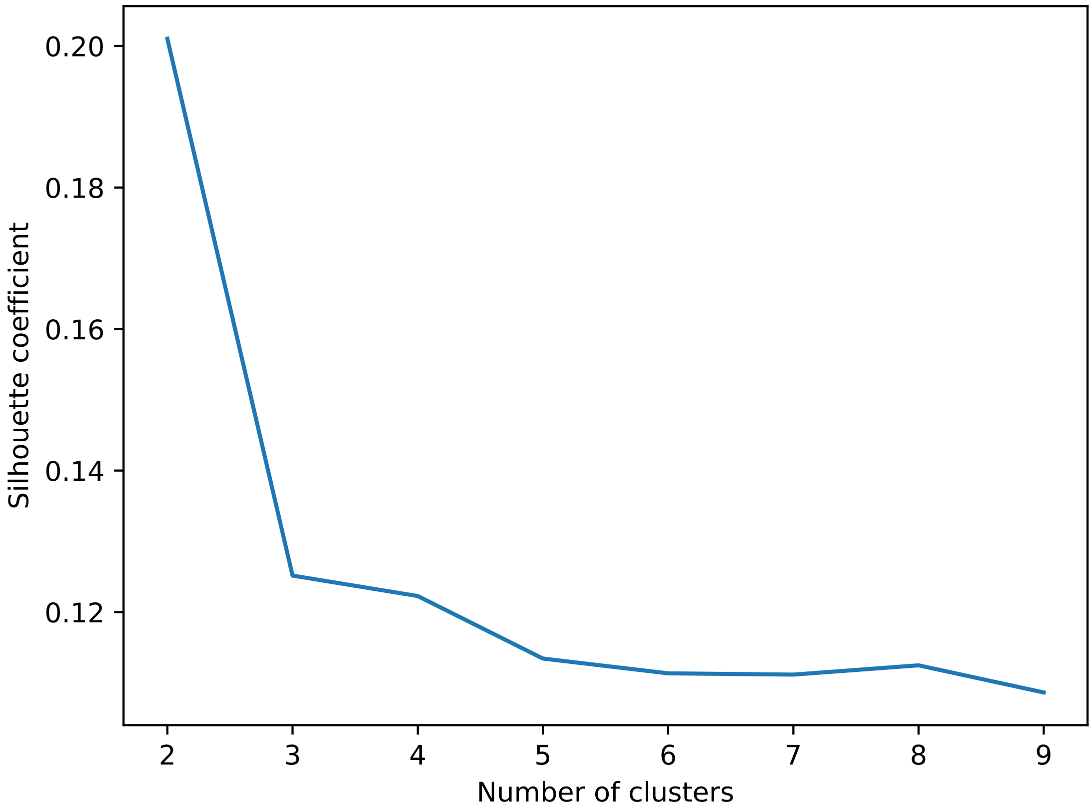

Figure 3: Silhouette plot for determining the optimal number of clusters.

Table 5
Comparison of results of evaluation metrics for K-Means clustering with and without PCA.

Metric	With PCA	Without PCA
Silhouette Coefficient ↑	0.43	0.21
Calinski-Harabasz Index ↑	258.27	88.82
Davies-Bouldin Index ↓	0.87	1.71

We use evaluation metrics such as Calinski-Harabasz Index [32] and Davies-Bouldin Index [33] to choose the one with best performance. In table 5, we can see that applying PCA gives us a comparatively a better evaluation score.

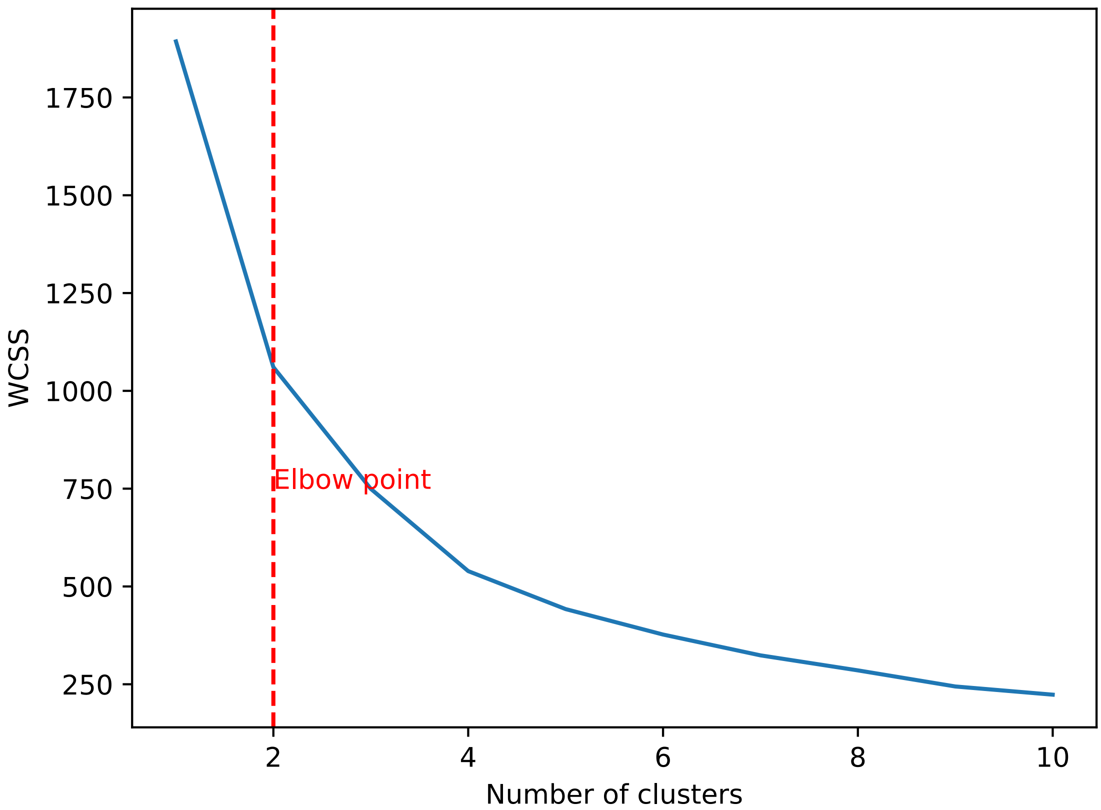

Figure 4: Elbow method for determining the optimal number of clusters with PCA.

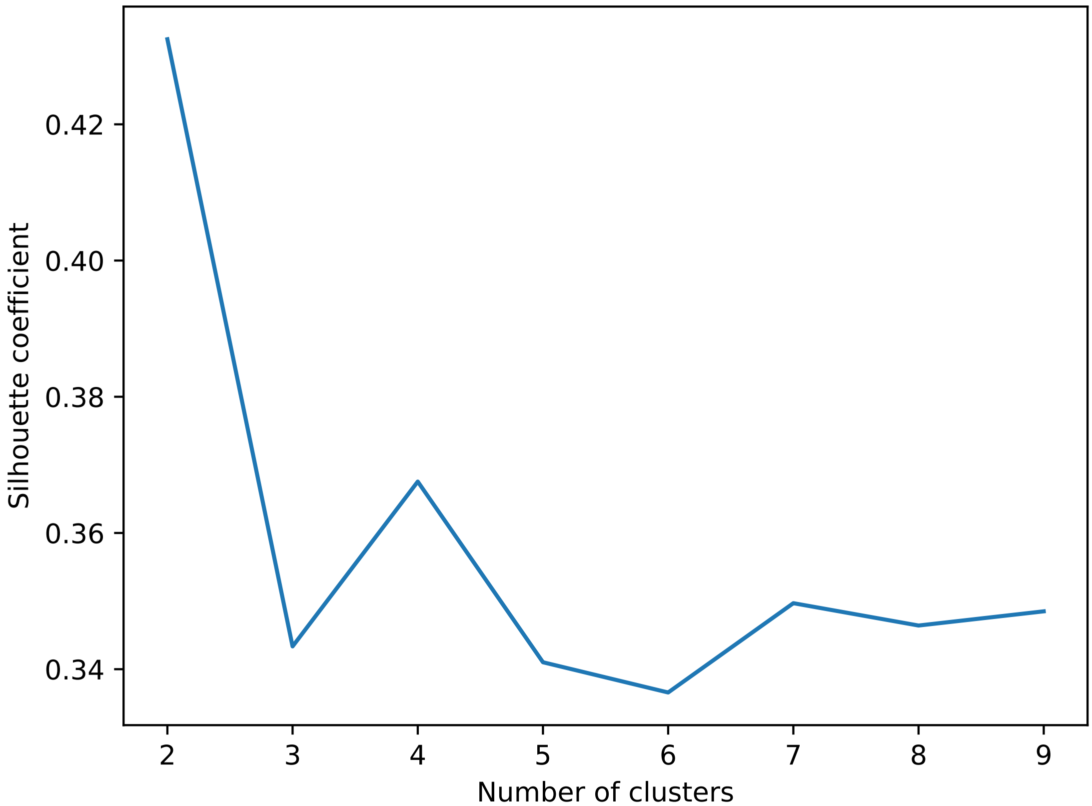

Figure 5: Silhouette plot for determining the optimal number of clusters with PCA.

4. Empirical Analysis

Now, we use the 331 valid responses out of the 343 responses we receive, to show some statistical description of the dataset and do some exploratory data analysis:

Based on our survey we see almost 331 people submitted their responses and most of the participants are male (61.03%). We have received 125 responses from female participants (37.76%) and also received 4 more responses from others (1.21%).

Analyzing the frequency distribution of BMI, most of the participants (58.3%) respond to having a normal BMI where the submission number is 193. The second largest submitted response (25.7%) for BMI is recorded to be overweight with 85 responses. Thirdly, almost 28 participants consider BMI as obese (8.5%). Moreover, 25 participants responded assuming to be underweight (7.6%).

Figure 6: Distribution of Online Food Orders Per Month.

From Figure 6 we can see the number of ordering food online per month by our survey participants. Here the range of ordering online is from 0 to 17. The highest number of ordering food in a month is 3 according to the chart. After that 2 and 1 are respectively the second and third highest numbers.

Our online food ordering survey participants are mostly undergraduate-level students (76.13%) where the number is 252. We also have 57 responses from Graduate / Post-Graduate / Ph.D. students (17.22%). The least amount of responses (6.65%) we received from higher school certificate (HSC) / A-Level students with 22 responses.

From table 9, we see that 146 participants are seeking career opportunities (44.11%). We also have 62 participants who are full-time employed (18.73%). A significant number of participants (17.52%) did not mention their employment status where the number is 58. Moreover, 65 participants are doing part-time jobs (19.64%).

Inspecting the table 10, we see that 142 participants are fully dependent (42.90%). The number of partially dependent (41.09%) participants is 136. Lastly, 53 people responded to their status as independent (16.01%).

On table 11, we can see that 304 individuals are single (91. 84%). We also have 18 participants who are married (5.44%) and only 9 respondents did not mention their marital status (2.72%).

The figure 7 contains the distribution of the respondents' physical activity over a week. More than 160 participants responded that not performing any physical activity per week. Less than 60 participants responded that they are doing physical activities 2 to 3 times per week. Less than 40 participants voted 1 time engaging in physical activities and not more than 20 people responded 4 to 7 times per week of investing time per week for physical activities.

Table 6
Distribution of Gender.

Gender	Frequency	Percentage
Male	202	61.03%
Female	125	37.76%
Prefer not to say	4	1.21%

Table 7
Distribution of BMI.

BMI	Frequency	Percentage
Underweight	25	7.6%
Normal	193	58.3%
Overweight	85	25.7%
Obese	28	8.5%

Table 8
Distribution of Educational Qualification.

Education	Frequency	Percentage
Undergraduate	252	76.13%
High School	22	6.65%
Graduate/Post-Graduate/Phd	57	17.22%

Table 9
Distribution of Employment Status.

Employment Status	Frequency	Percentage
Seeking opportunities currently	146	44.11%
Part-time	65	19.64%
Full-time	62	18.73%
Prefer not to say	58	17.52%

We can see that 262 participants are ordering food online for more than 6 months (80.66%). Moreover, we can notice that 43 participants ordered food online for 3 months (12.99%). Furthermore, 21 participants are ordering food online for 3 to 6 months (6.34%).

Table 13 gives us an idea of when people tend to order food. Firstly, most of the participants (38.22%) are ordering food for evening snacks and the number is 227. Secondly, 176 participants voted for online food ordering time as dinner (29.63%). Thirdly, 138 people responded to lunchtime for ordering food online (23.23%). Moreover, The number of ordering food online for midnight snacks (6.06%) is 36. Furthermore, 17 people selected breakfast for ordering food online (2.86%).

Most of the survey participants (59.57%) are using 'foodpanda' and the number is 302. The second highest (26.82%) used app by our survey participants is 'Pathao' where the number of users is 136. Thirdly, 55 people are using 'Hungrynaki' (10.85%). Lastly, we can see that 14 survey participants are using other apps (2.76%).

From figure 8, we can see that majority tend to order burgers (267) and pizzas (249) when ordering food online, which consist of high amount of saturated fat and salt. Salads (19) is among the lowest to be ordered. There were some food types with only one entry, we dropped those since those are negligible.

The figure 9 gives us an idea of the importance of each factors on a Likert scale. Firstly, we found that 100 people agreed and 125 responded as neutral to being inclined to trying out new technologies. Secondly, 120 people voted neutral and 73 agreed to making impulsive decisions to order food online. Thirdly, 97 people responded neutral, 86 agreed and 77 disagreed with being too busy to cook food themselves and leaning towards OFO services. There is

Table 10
Distribution of Financial Dependency.

Financial Dependency	Frequency	Percentage
Fully Dependent	142	42.90%
Partially Dependent	136	41.09%
Independent	53	16.01%

Table 11
Distribution of Marital Status.

Marital Status	Frequency	Percentage
Single	304	91.84%
Married	18	5.44%
Prefer not to say	9	2.72%

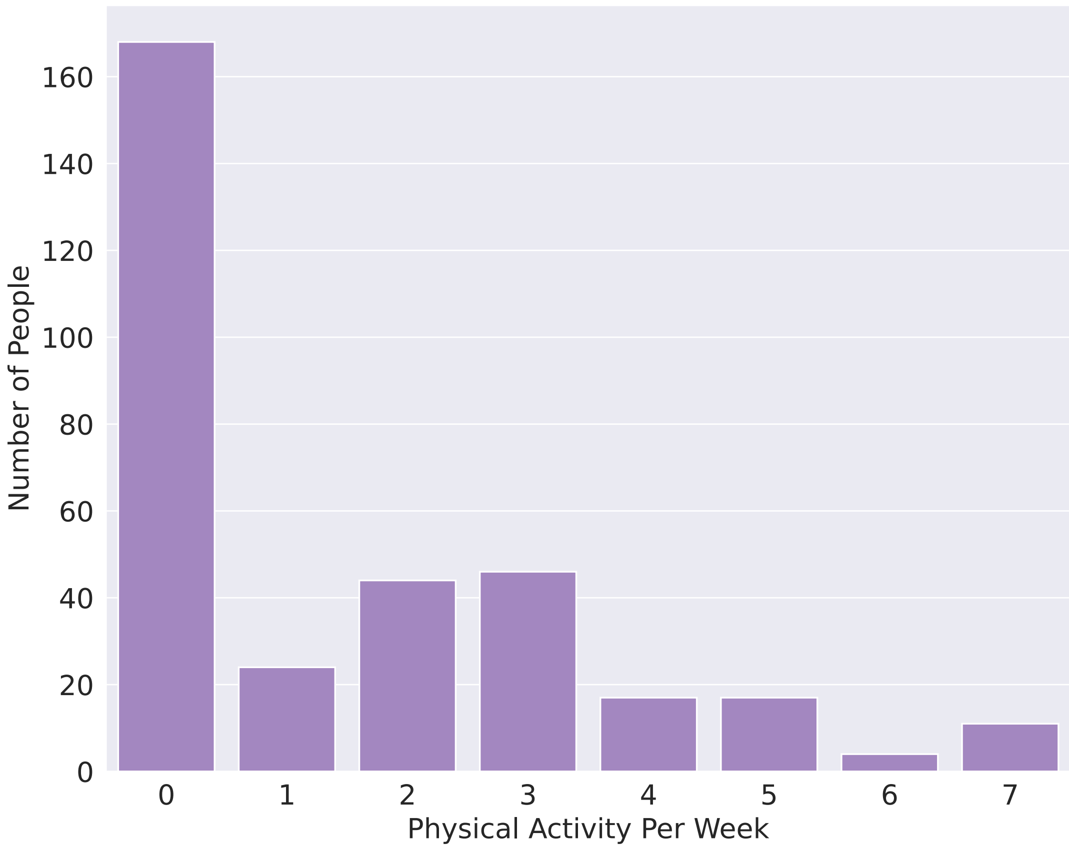

Figure 7: Distribution of Physical Activity.

a match with 83 numbers of responses where each is allocated for disagree and neutral where the factor is people don't like to cook. A huge number of participants agreed that ordering food online is easy, where 152 people agreed and 102 strongly agreed. Again, lots of people agreed that they like having variety of options to choose their desired food from OFO platforms, where 139 people agreed and 114 strongly agreed. In the case of choosing to order food online due to being inexpensive, 110 participants responded as neutral and 98 people disagreed. 117 people agreed that

Table 12
Distribution of Online Food Ordering History.

Ordering History	Frequency	Percentage
Less than 3 months	43	12.99%
3-6 months	21	6.34%
More than 6 months	262	80.66%

Table 13
Distribution of Online Food Ordering Time.

Ordering Time	Frequency	Percentage
Breakfast	17	2.86%
Lunch	138	23.23%
Evening Snacks	227	38.22%
Dinner	176	29.63%
Midnight Snacks	36	6.06%

Table 14
Distribution of the Use of Online Food Ordering Apps.

OFO Apps	Frequency	Percentage
foodpanda	302	59.57%
HungryNaki	55	10.85%
Pathao	136	26.82%
Others	14	2.76%

promotional discounts convince them to order food online whereas 86 respond to neutral. A big portion of our survey participants agreed that ordering food online allows them to get their food quickly and the response number is 120, besides 116 people voted neutral. 174 of the participants stay neutral when they are asked if they think their food is safe and hygienic and 78 participants disagreed with it. Lastly, a very few survey participants believe that their online food ordered item is nutritious with 118 people responding as neutral, 106 disagreed and 63 participants strongly disagreed with the factor.

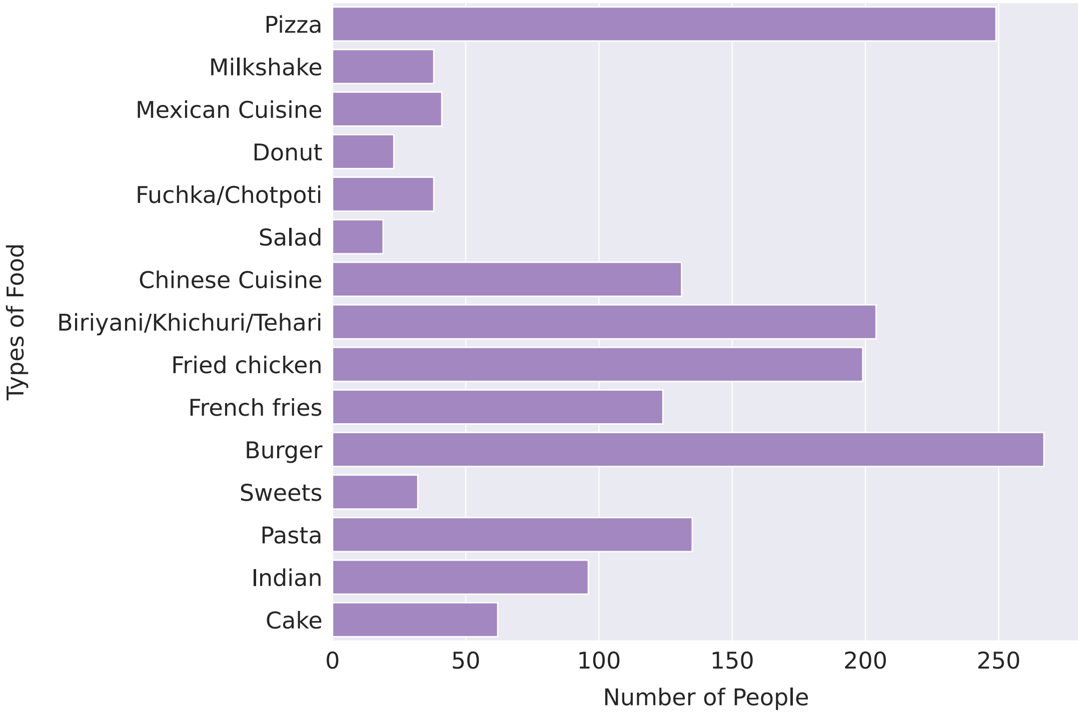

Figure 8: Distribution of type of foods ordered through online platforms.

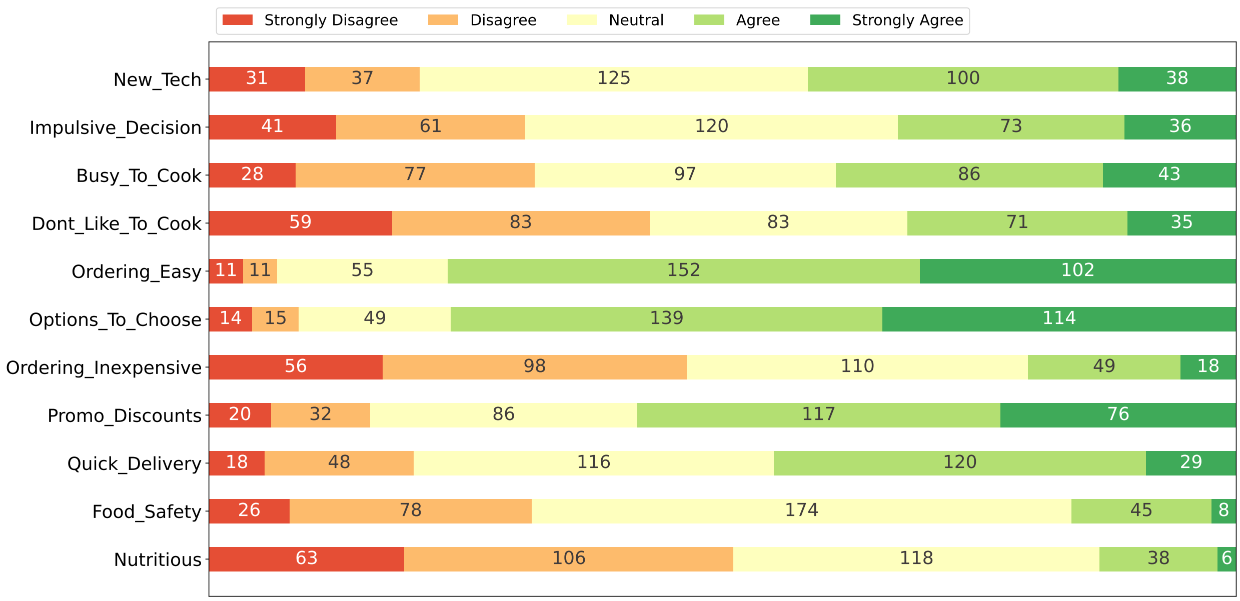

Figure 9: Distribution of Factors that Influence Online Food Orders.

5. Results

The results of the experiments which are done to find the most influencing factors are shown in table 15 gives us an idea on which factors influence a person to use OFO services.

Table 15
Influential Factors in Online Food Orders: Comparing Mean, Standard Deviation, OLS, Decision Tree, and Random Forest Values.

Factors	Mean	SD	OLS	DT	RF
New Tech	2.23	1.09	0.39	0.01	0.22
Impulsive Decision	2.01	1.16	0.90	0.34	0.22
Busy To Cook	2.12	1.16	0.18	0.05	0.10
Don't Like To Cook	1.82	1.25	0.33	0.01	0.13
Ordering Easy	2.98	0.95	0.65	0.11	0.08
Options To Choose	2.98	1.03	0.15	0.01	0.06
Ordering Inexpensive	1.62	1.10	0.27	0.22	0.07
Promo Discounts	2.60	1.12	0.33	0.01	0.09
Quick Delivery	2.28	1.00	0.11	0.01	0.05
Food Safety	1.79	0.86	-0.40	0.01	0.05
Nutritious	1.45	0.98	-0.32	0.28	0.09

Table 16
Results of Point-biserial Correlation on BMI vs Orders Per Month.

BMI	P-value	Coefficient	Relationship
Underweight	0.84	-0.01	Rejected
Normal	1.97^{-5}	-0.23	Weak Negative
Overweight	0.04	0.04	Very Weak Positive
Obese	1.13^{-10}	0.34	Weak Positive

The mean values based on solely the responses from the respondents suggests that the most important factors in influencing people's decision to order food online are the convenience of ordering (mean score of 2.98), the variety of options available (mean score of 2.98), and promotional offers and discounts (mean score of 2.60).

The results of the hypothesis study point out that the first hypothesis, H1 is accepted and the null hypothesis is rejected as the p-value is < 0.001, which is less than ($\alpha = 0.05$). This means that people who tend to try out new technologies also feel that ordering food online is really convenient and easy.

For the second hypothesis, H2 is rejected as the p-value (0.1143) is higher than the α (0.05). This means we could not find any significant relationship between being too busy to cook and making impulsive decisions when ordering food online.

The third hypothesis, H3 is accepted and the null hypothesis is rejected as the p-value (< 0.001) is lower than the α (0.05). This means there is a significant relationship between having many options to choose from when ordering food online and not liking to cook.

And the fourth hypothesis, H4 is accepted and the null hypothesis is rejected as the p-value (< 0.001) is less than the α (0.05). This means that there exists a significant relationship between finding ordering food online to be inexpensive and using promo codes and discounts when ordering food online.

The results from Pearson correlation and Spearman's rank correlation which was done between BMI and orders per month, gives us p-value of 2.43^{-7} and 0.003 respectively, which suggests that there exists a statistically significant positive correlation between the two variables since both these values are less than 0.05. The coefficients values are 0.28 and 0.15 respectively which suggests there exists a weak positive correlation and a very weak positive correlation [34].

The results of shown on table 16 suggests that due to insufficient p-value any relation between Underweight and orders per month is rejected, normal weight has a weak negative relationship, overweight has a very weak positive relationship and obese has a weak positive relationship.

As seen on table 17, we can see that there exists a very weak negative relation between physical activity and orders per month which suggests that people who tend to order more may exercise less. However, the results from Spearman's test is rejected as the p-value is 0.085, which is greater than the α value of 0.05.

The results of correlation analysis between ordering history and orders per month as seen on table 18 shows that people who have been using OFO for more than 6 months tend to order more. The coefficient value of -0.23 suggests that there exists a weak relationship between ordering frequency and people who have been using this type of service

Table 17
Results of Correlation analysis between physical activity and orders per month.

Test	P-value	Coefficient	Relationship
Pearson	0.026	-0.12	Very Weak Negative
Spearman	0.085	-0.10	Rejected

Table 18
Results of Point-biserial Correlation on Ordering History and Orders Per Month.

Ordering History	P-value	Coefficient	Relationship
$<$ 3 months	1.40^{-5}	-0.23	Weak Negative
3 - 6 months	0.66	0.02	Rejected
$>$ 6 months	<0.0001	0.18	Very Weak Positive

Table 19
Comparison of results among prediction models before feature selection.

Model	Accuracy	Precision	Recall	F1-Score
LR	0.78	0.66	0.59	0.61
NB	0.64	0.53	0.53	0.53
DT	0.63	0.45	0.45	0.45
RF	0.81	0.78	0.59	0.60
GBM	0.78	0.66	0.59	0.61
KNN	0.78	0.67	0.64	0.65

Table 20
Comparison of results among prediction models after feature selection.

Model	Accuracy	Precision	Recall	F1-Score
LR	0.81	0.74	0.61	0.63
NB	0.67	0.55	0.55	0.55
DT	0.61	0.49	0.49	0.49
RF	0.81	0.78	0.59	0.60
GBM	0.78	0.65	0.57	0.58
KNN	0.81	0.73	0.64	0.66

for less than 3 months. However, due to the p-value (0.66) being higher than the α value of 0.05, any relationship between people who have been ordering for 3-6 months is rejected.

The results of the prediction model can be seen on table 19.

We can see that out of 6 models, Random Forest performs the best with 81% accuracy. whereas Naive Bayes performed worst with 64% accuracy. Random Forest also has the best precision score of 0.78 which means it has a high level of precision in identifying positive instances. On the other hand, KNN correctly identified 64% of the positive instances as seen on the Recall column.

On table 21, we have the mean values of each cluster. We show the values of both with PCA applied and without it. On cluster 0 which contains respondents who tend to order less and cluster 1 represents the people who order comparatively more. We can see that people of cluster 0 find food delivered by OFO services to not be inexpensive with the mean value for 'Ordering Inexpensive' as 0.83 and 0.97. Whereas, people belonging in cluster 1 tend to find the food from the OFO services to lack nutritious values with mean value for 'Nutritious' as 1.64 and 1.66.

Table 21
Comparison between cluster means between the clusters created with and without PCA. Here the factors are New Tech (NT), Impulsive Decision (ID), Too Busy To Cook (BC), Don't Like to Cook (DLC), Ordering is Easy (OE), Options to Choose (OC), Ordering is Inexpensive (OI), Promotional Offers and Discounts (POD), Quick Delivery (QD), Food Safety (FS), Food is Nutritious (FN).

Factors	With PCA		Without PCA	
	Cluster 0	Cluster 1	Cluster 0	Cluster 1
NT	1.45	2.47	1.45	2.48
ID	1.19	2.25	1.19	2.27
BC	1.47	2.31	1.47	2.34
DLC	1.08	2.04	1.08	2.04
OE	1.95	3.29	1.95	3.38
OC	1.75	3.35	1.75	3.44
OI	0.83	1.86	0.83	1.90
POD	1.53	2.92	1.53	3.02
QD	1.29	2.59	1.29	2.65
FS	1.17	1.98	1.17	2.02
FN	1.29	1.64	1.29	1.66

6. Discussions

The results of OLS show that the most influencing factors are making impulsive decision (coefficient value of 0.90), convenience of ordering (coefficient value of 0.65). On the other hand, food safety (coefficient value of -0.40) and nutritious value of the food (coefficient value of -0.36) have a negative impact. The results of the most influential factor mining supports some of the findings from previous works. The high coefficient value on impulsive decision supports [4] as they also find that hedonic motivation is one of the important factors. The positive impact of ease of convenience as found by [3, 4] is supported by our results as well. We found that food safety is a major concern as it has a negative impact when using OFO services, this is also supported by [11].

Additionally, we can see that the results of the hypothesis testing shows that hypotheses, H1, H2, H3 were accepted and H2 was rejected. This suggests some of the factors are not necessarily independent of each other.

The results of Pearson Correlation suggests that when we treat BMI as a continuous value, there exists a weak positive correlation between BMI and orders per month. Although this does not indicate that, there is necessarily any causation here. When treating BMI as categorical value, the results of Spearman Correlation show that there exists a weak positive correlation. Although this does not indicate that, there is necessarily any causation here. Moreover, when we treat the BMI values as binary categories it shows that there exists a very weak positive and a weak positive correlation for overweight and obese respectively. However, people with normal BMI show weak negative relationship with orders of food per month. So, even though there exists a positive relationship between being overweight or obese and ordering a lot of food per month, it is however very weak. These findings can be used in future studies to combine with other factors to see if any causality actually exists.

We can see on table 20, that after applying feature selection the performance of Logistic Regression, Naive Bayes and KNN improves. Although, performance of Decision Tree decreases and that of Random Forest and Gradient Boosting Machine stays the same as before. This can also be seen on figure 10.

Moreover, the results of customer segmentation show that respondents in both clusters show the tendency to order food because it is convenient, there are variety of options to choose from and due to the availability of promotional offers and discounts. However, we can see that people of cluster 0 find food delivered by OFO services to not be inexpensive. Whereas, people belonging in cluster 1 tend to find the food from the OFO services to lack nutritious values.

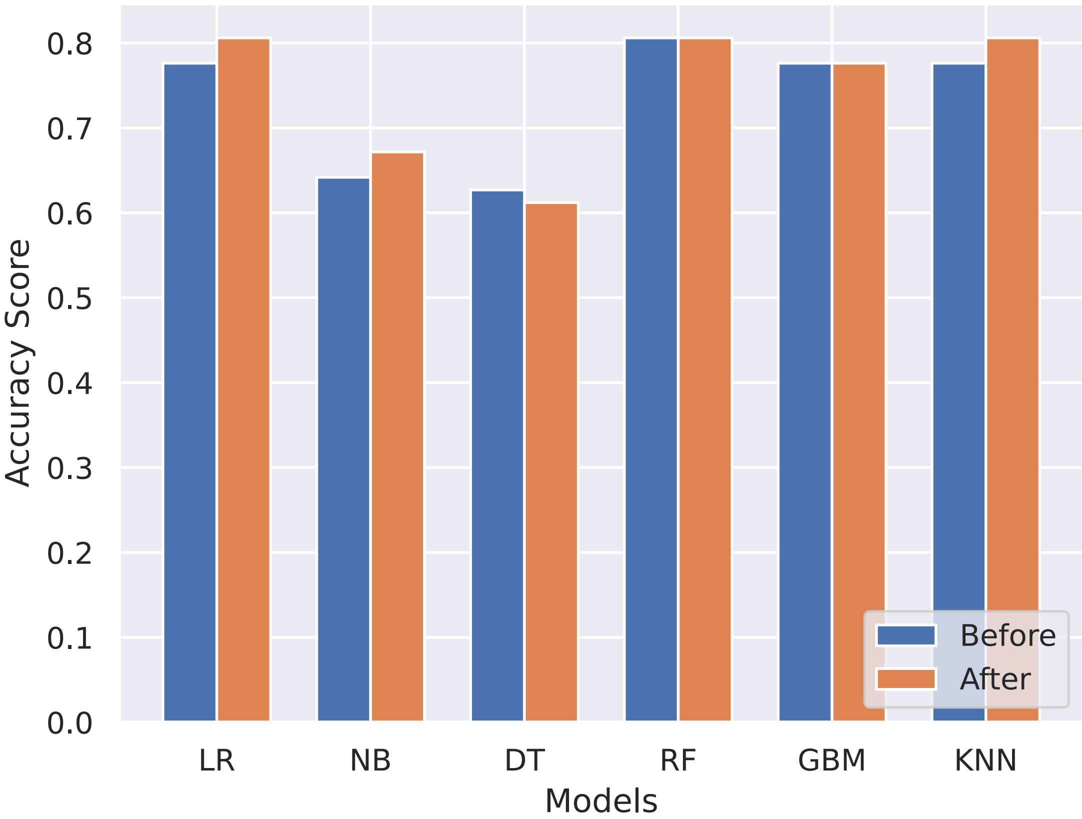

Figure 10: Comparison between results of prediction models before and after applying feature selection.

7. Conclusion

This research tried to investigate the factors that influence customers to use online food ordering platforms in the context of Bangladesh. We were able to create a dataset containing the activities of OFO service users, which is non existent specially in Bangladeshi context. Our findings reveal that there exists a weak positive correlation between BMI and orders per month. It also shows that people who tend to order more, happen to be obese. However, it is important to note that a correlation does not necessarily imply causation. We found that making impulsive decision, convenience of ordering and the availability of promotional offers and discounts are the most important factors behind the frequency of ordering food online. We also were able to successfully classify frequent users of OFO services with 81% accuracy.

7.1. Limitations & Future Work

Our research has its limitations. First of all, our data was collected from people of ages 16-30, which means that it may not accurately reflect the characteristics and perspectives of the entire population. Moreover, our study was cross-sectional, which means that we cannot make causal inferences from our results. Future research could be done with even bigger data with more factors to consider. Another research opportunity is to include more factors such as advertisements, spending level to get better insights.

CRediT authorship contribution statement

S. M. Fazle Rabby Labib: Conceptualization, Formal Analysis, Investigation, Methodology, Software Visualization, Writing - original draft. **Fahmida Zaman Achol:** Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing - original draft. **Md. Aqib Jawwad:** Conceptualization, Formal Analysis, Investigation, Methodology, Writing - original draft. **Md. Golam Rabiul Alam:** Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Ashis Talukder:** Funding acquisition.

References

[1] W. H. Organization, Obesity and overweight, https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight, 2021.

[2] NHS, Obesity causes, https://www.nhs.uk/conditions/obesity/causes/, 2023.

[3] G. Kale, S. Chourishi, The customer perception on online food ordering and its significance, Alochana Chakra Journal (2020).

[4] S. M. Alagoz, H. Hekimoglu, A study on tam: Analysis of customer attitudes in online food ordering system, Procedia - Social and Behavioral Sciences 62 (2012) 1138–1143.

[5] R. Parameshwaran, K. Krishnasamy, A study on factors influencing online consumer buying behavior, International Journal of Scientific & Technology Research 9 (2020) 3620–3625.

[6] S. Anam, D. Golam Yazdani Showrav, M. A. Hassan, A. Chakrabarty, Factors influencing the rapid growth of online shopping during covid-19 pandemic time in dhaka city, bangladesh, Academy of Strategic and Management Journal 20 (2021) 1–13.

[7] A. Chatterjee, M. W. Gerdes, S. G. Martinez, Identification of risk factors associated with obesity and overweight—a machine learning overview, Sensors 20 (2020) 2734.

[8] A. M. Prentice, S. A. Jebb, Fast foods, energy density and obesity: a possible mechanistic link, Obesity Reviews 4 (2003) 187–194.

[9] L. A. H. Harahap, E. Aritonang, Z. Lubis, The relationship between type and frequency of online food ordering with obesity in students of medan area university, Britain International of Exact Sciences (BIoEx) Journal 2 (2020) 29–34.

[10] N. D. Kurniawati, S. N. Cahyaningsih, A. S. Wahyudi, The correlation between online food ordering and nutritional status among college students in surabaya, Indonesian Journal of Community Health Nursing 6 (2021) 70.

[11] C. Hong, H. H. Choi, E.-K. C. Choi, H.-W. D. Joung, Factors affecting customer intention to use online food delivery services before and during the COVID-19 pandemic, Journal of Hospitality and Tourism Management 48 (2021) 509–518.

[12] E. Martha, D. Ayubi, B. Besral, N. D. Rahmawati, A. P. Mayangsari, Y. Sopamena, M. Astari, R. S. Zulfa, Online food delivery services among young adults in depok: Factors affecting the frequency of online food ordering and consumption of high-risk food, Mapping Intimacies (2021).

[13] I. Jahidi, N. A. Ruyani, D. P. Alamsyah, The study of consumer behavior on online food ordering system (go-food) in the metropolitan city (2022).

[14] C. Inthong, T. Champahom, S. Jomnonkwao, V. Chatpattananan, V. Ratanavaraha, Exploring factors affecting consumer behavioral intentions toward online food ordering in thailand, Sustainability 14 (2022) 8493.

[15] K. S. Taber, The use of cronbach's alpha when developing and reporting research instruments in science education, Research in Science Education 48 (2017) 1273–1296.

[16] B. Zdaniuk, Ordinary Least-Squares (OLS) Model, Springer Netherlands, Dordrecht, 2014, pp. 4515–4517. URL: https://doi.org/10.1007/978-94-007-0753-5_2008. doi:10.1007/978-94-007-0753-5_2008.

[17] J. Fürnkranz, Decision Tree, Springer US, Boston, MA, 2010, pp. 263–267. URL: https://doi.org/10.1007/978-0-387-30164-8_204. doi:10.1007/978-0-387-30164-8_204.

[18] T. K. Ho, Random decision forests, in: Proceedings of 3rd international conference on document analysis and recognition, volume 1, IEEE, 1995, pp. 278–282.

[19] K. Pearson, X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50 (1900) 157–175.

[20] D. Freedman, R. Pisani, R. Purves, Statistics (international student edition), Pisani, R. Purves, 4th edn. WW Norton & Company, New York (2007).

[21] C. Spearman, Spearman Rank Correlation Coefficient, Springer New York, New York, NY, 2008, pp. 502–505. URL: https://doi.org/10.1007/978-0-387-32833-1_379. doi:10.1007/978-0-387-32833-1_379.

[22] L. J.M, The expected value of a point-biserial (or similar) correlation, 2008. URL: https://www.rasch.org/rmt/rmt221e.htm.

[23] D. R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society: Series B (Methodological) 20 (1958) 215–232.

[24] G. I. Webb, Naïve Bayes, Springer US, Boston, MA, 2010, pp. 713–714. URL: https://doi.org/10.1007/978-0-387-30164-8_576. doi:10.1007/978-0-387-30164-8_576.

[25] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[26] A. Mucherino, P. J. Papajorgji, P. M. Pardalos, k-Nearest Neighbor Classification, Springer New York, New York, NY, 2009, pp. 83–106. URL: https://doi.org/10.1007/978-0-387-88615-2_4. doi:10.1007/978-0-387-88615-2_4.

[27] X. Jin, J. Han, K-Means Clustering, Springer US, Boston, MA, 2010, pp. 563–564. URL: https://doi.org/10.1007/978-0-387-30164-8_425. doi:10.1007/978-0-387-30164-8_425.

[28] I. Jolliffe, Principal component analysis, in: International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, 2011, pp. 1094–1096. URL: https://doi.org/10.1007/978-3-642-04898-2_455. doi:10.1007/978-3-642-04898-2_455.

[29] H. F. Kaiser, A second generation little jiffy, Psychometrika 35 (1970) 401–415.

[30] M. S. Bartlett, Tests of significance in factor analysis., British journal of psychology (1950).

[31] A. Zbiciak, T. Markiewicz, A new extraordinary means of appeal in the polish criminal procedure: the basic principles of a fair trial and a complaint against a cassatory judgment, Access to Justice in Eastern Europe 6 (2023) 1–18.

[32] C. T. Harabasz, M. Karoński, A dendrite method for cluster analysis, in: Communications in Statistics, volume 3, 1974, pp. 1–27.

[33] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE transactions on pattern analysis and machine intelligence (1979) 224–227.

[34] Z. Jaadi, Eveything you need to know about interpreting correlations, 2019. URL: https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8.