CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The introduction of the thesis is covered in this part. The main focus of the thesis study is the COVID-19's impacts analysis on business operations and earnings. Therefore, the background study's initial emphasis is on the problems with the COVID-19 effects analysis. Furthermore, a history with COVID-19 effects research is required. The problem statement is then presented, after which the research objectives are created based on it. The thesis's scope is provided after this chapter.

1.2 BACKGROUND STUDY

The extremely infectious and potentially fatal respiratory disease COVID-19 is brought on by a new corona virus (SARS-CoV-2). In December 2019, the virus was discovered in Wuhan, China. Since then, it has spread quickly over the world, causing a pandemic. The COVID-19 symptoms might be minor to severe, and other individuals report having no symptoms at all. Fever, coughing, and breathing issues are typical symptoms. When an infected individual coughs, sneezes, or speaks, the virus is mostly transmitted by respiratory droplets. Governments and health groups have employed a few tactics to stop the virus's transmission, including lockdowns, social isolation, and mask regulations. Vaccines have also been developed and are being distributed internationally to help curb the virus's spread and protect those who are vulnerable.

The COVID-19 epidemic has a significant and unprecedented impact on the world economy, notably hurting small enterprises. Several small and medium-sized businesses have battled to stay open in the aftermath of the epidemic; some have even had to close permanently. Pre-impacts, mid-impacts, and post-impacts phases, which denote the distinct stages of the pandemic's progression, have all had varied effects on nearby companies at various times.

The pre-impacts phase refers to the period before the pandemic had a significant impact on the economy, when businesses were operating as usual. During this phase, businesses were preparing for the pandemic’s arrival and trying to anticipate its impact on their operations. In the mid-impacts phase, businesses began to feel the effects of the pandemic as government-mandated restrictions and social distancing measures were forced to shut down or reduce their operations, resulting in significant losses. Finally, the post-impacts phase refers to the period after the pandemic, where businesses were trying to recover and adapt to the new normal.

This study intends to look at how COVID-19 affected neighborhood businesses at each of these stages. This thesis will examine the issues encountered by several local business kinds, such as small food bakeries, meat stores, and poultry business, via a thorough literature research and case study analysis. Throughout each stage of the epidemic, companies employed various adaptation and survival methods, which will also be looked at in the research. This study aims to shed light on how neighborhood businesses have responded to the COVID-19 outbreak and to pinpoint the best methods for assisting them in emergency situations. This study attempts to give a thorough knowledge of the pandemic's effects on Businesses by examining their pre-, mid-, and post-impact stages of the COVID-19 pandemic's difficulties and prospects for these companies. In the end, this study will add to the body of knowledge about how COVID-19 has affected neighborhood businesses and offer insightful information to both politicians and company owners.

1.3 PROBLEM STATEMENT

Businesses across the world have been significantly impacted by the Covid-19 outbreak. Local firms have had to adjust to shifting customer habits, new rules from the government, and problems with the supply chain. As a result, many companies have seen declines in revenue and profitability as well as closures. The goal is to comprehend how much the Covid-19 has affected nearby businesses and to find solutions that would enable them to bounce back and adjust to the new normal.

Businesses are still facing many difficulties as a result of the ongoing Covid-19 outbreak, including decreased foot traffic, supply chain interruptions, elevated operational expenses, and shifting consumer preferences. As a result, many neighborhood businesses are having a difficult time surviving, and some even face closing. The goal of the issue statement is to identify the present problems that Businesses are facing, come up with solutions for them, and create plans that will enable them to survive the changing business environment.

A prosperous, sustainable, and resilient future is what Businesses aim for. Businesses must be able to adjust to shifting consumer demands and technological developments while still taking care of social and environmental issues. In order to build a conducive ecosystem for small enterprises, this future should also involve increasing cooperation between neighborhood businesses, local governments, and other stakeholders. Local firms should also be able to make use of digital tools and platforms to increase their reach and take on bigger rivals. In the end, Businesses hope for a future where they can generate significant economic opportunities, encourage community involvement, and help to build a more just and successful society.

Considering the substantial quantity of study done on the effects of Covid-19 on neighborhood businesses, numerous research gaps still need to be filled. Understanding the diverse effects of the pandemic on various local business types, such as those in rural vs. urban locations or those in various industries, is one study gap. The need to examine the pandemic's longer-term effects on local firms, such as the degree to which consumer behavior shifts and supply chain disruptions are expected to endure even after the virus has passed, represents another study need. The success of various policy initiatives, such as grants, loans, and tax incentives, targeted at assisting local enterprises must also be investigated.

Here we noted a list of problem from previous studies:

(a) Many of the studies were conducted during the early stages of the pandemic, and thus may not reflect the full extent of the impact on businesses over a longer time frame.

(b) Some studies focused on a specific region or industry, and thus the findings may not be representative of the wider business landscape.

(c) Some studies focused on specific aspects of the impact of COVID-19 on businesses, such as financial health or supply chains, and may not provide a comprehensive view of the overall impact.

(d) The studies focused on specific regions or industries and may not be applicable to other regions or industries.

(e) Some studies relied on self-reported data from businesses, which could be subject to bias or inaccuracies.

(f) Many research focuses mainly on the impact of COVID-19 on businesses in developed countries, with little attention given to developing nations.

(g) Most of the research focuses on the negative impacts of COVID-19 on businesses, and there is a need for more research on potential opportunities and positive outcomes.

(h) The majority of them used a small number of Machine learning algorithms as a prediction algorithm.

(i) Source of data are very limited.

1.4 RESEARCH OBJECTIVE

The goal of this study is to use statistical analysis and machine learning algorithms to examine the effects of COVID-19 on Businesses. The following sub-objectives must be realized in order to do this:

I. To identify the pre-impact factors that are most predictive of business performance during the COVID-19 pandemic, including number of employees, business time, sales amount, expenditure, and profit.

II. To evaluate the mid-impact effects of COVID-19 on businesses, including changes in employee numbers, sales amounts, expenditure, and profits.

III. To assess the post-impact effects of COVID-19 on businesses, including recovery rates, changes in consumer behavior, and long-term impacts on business performance.

IV. To compare the performance of linear regression, random forest, and KNN models for predicting business profits on the COVID-19 pandemic.

V. To develop a prediction model that accurately forecasts business profits on the COVID-19 pandemic, using a combination of pre-impact, mid-impact, and post-impact factors.

VI. To provide recommendations for businesses based on the findings of the study, including strategies for adapting to the challenges presented by the pandemic and optimizing business performance in the future.

1.5 SCOPE OF THE RESEARCH

The scope of our research on the pre-impacts, mid-impacts, and post-impacts of COVID-19 on businesses can be quite broad, as it covers a significant and ongoing event that has affected businesses across the world. Here are some aspects that we may consider within the scope of our research:

a) Geographical scope: Our research can cover businesses in specific countries or regions, or it can take a global perspective.
b) Industry scope: We can focus on specific industries that have been most impacted by COVID-19, such as tourism, hospitality, or retail, or consider a broader range of industries.
c) Time frame: We can specify the time frame for our research, such as the first wave of the pandemic, or a longer period of time that covers multiple waves.
d) Size of businesses: We can consider businesses of different sizes, from small businesses to large corporations, to examine how COVID-19 has impacted businesses of different scales.
e) Type of analysis: Our research can involve both quantitative and qualitative analysis, with the former focusing on statistical analysis and prediction models, while the latter may involve case studies, interviews, and surveys.

.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The purpose of this chapter is to provide a review of past research efforts related to COVID-19 impacts analysis using various methods. This chapter also shows the summary of the related works.

2.2 RELATED WORK

Several relevant studies are discussed below:

A prediction model is proposed by Fatemeh Safara to anticipate the consumers behaviour using machine learning methods. Five individual classifiers, and their ensembles with Bagging and Boosting are examined on the dataset collected from an online business ping site. The results indicate the model constructed using decision tree ensembles with Bagging achieved the best prediction of consumer behavior with the accuracy of 95.3%[1] .

Another research is done by Jim Samuel and Yana Samuel which provides the insights of Corona-virus fear sentiment progression. They outlines associated methods, implications, limitations and opportunities.They observe that strong classification accuracy of 91% for short Tweets, with the Naive Bayes method and the logistic regression classification method provides a reasonable accuracy of 74% with shorter Tweets, and both methods showed relatively weaker performance for longer Tweets[2] .

Yi Luo and Xiaowei Xu show that deep learning algorithms like Bidirectional LSTM and Simple Embedding Average Pooling to outperform traditional machine learning algorithms in sentiment classification and review rating prediction. This study strengthens the extant literature by empirically analyzing restaurant reviews posted during the COVID-19 pandemic and discovering suitable deep learning algorithms for different text mining tasks[3].

Chinmay Chakraborty & Senthilkumar Mohan make a study with different machine-learning algorithms to analyse the COVID-19 health analysis and prediction. They use open data resources from Mexico and Brazil. They have an accuracy of 93% and the perceived mean of recall and the precision F1 Score of 0.79 on the dataset used for the model of Mexico and the model of Brazil has an accuracy of 69% and F1 score of 0.75 on the dataset studied[4].

Grigore Belostecinic and Radu Ioan Mogoș make study on economic and social impact during COVID-19 Pandemic. They use a mixed methodology, combining results obtained through a survey based on a self-administered electronic questionnaire, with a data mining analysis. The

expansion of telework in various economic areas is a result of adaptation to the new economic and social conditions caused by the COVID-19 pandemic[5].

Charlyn Villavicencio and Julio Jerison Macrohon make a research with Natural language processing techniques were applied to understand the general sentiment, which can help the government in analyzing their response. The sentiments were annotated and trained using the Naive Bayes model to classify English and Filipino language tweets into positive, neutral, and negative polarities through the rapid miner data science software. The results yielded an 81.77% accuracy, which outweighs the accuracy of recent sentiment analysis studies using Twitter data from the Philippines[6] .

Georg Licht & Simona Murmann presented a statistical analysis to find out small firms and the COVID-19 insolvency gap. The insolvency gap is estimated to affect around 25,000 companies, a substantial number compared to the around 16,300 actual insolvencies in 2020. In the ongoing crisis, policy makers should prefer instruments favoring entrepreneurs who respond indicatively to the pandemic instead of prolonging the survival of near-insolvent firms[7] .

Rudy Arthur makes studies on the UK job market during the COVID-19 crisis with online job ads. An open methodology that enables a rapid and detailed survey of the job market in unsettled conditions and describes a web application jobtrender.com that allows others to query this data set[8].

Ting Zhang & Zoltan Acs built a theoretical framework based on firm profit maximization, compiled an up-to-date (March through November) real-time daily and weekly multifaceted data set, and empirically estimated fixed-effect panel data, fractional logit, and multilevel mixed effects models to test their hypotheses, find that in states with higher WFH rates, small businesses performed better overall with industry variations, controlling for the local pandemic, economic, demographic, and policy factors[9].

Andrius Grybauskas & Alina Stundžienė works on a predictive analytic using Big Data for the real estate market during the COVID-19 pandemic. 15 different machine learning models were applied to forecast apartment revisions, and the SHAP values for interpret ability were used. They find previous literature results, affirming that real estate is quite resilient to pandemics, as the price drops were not as dramatic as first believed[10].

2.3 SUMMARY OF THE RELETED WORK

Serial number	Title and Reference Number	Methodology	Contributions	Limitations
1	"A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic".[1]	Machine learning methods with five individual classifiers,	The best prediction of consumer behavior with the accuracy of 95.3%. In addition, correlation analysis is performed to determine the most important features.	Data limitations, Algorithm limitations, Generalizability limitations
2	"COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification".[2]	Machine learning (ML) classification methods, with the Naïve Bayes, logistic regression classification method.	This research provides insights into Coronavirus fear sentiment progression, and outlines associated methods, implications, limitations and opportunities.	Limited data availability, Uncertainty in predicting the future, Limited scope.
3	"Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic".[3]	Machine learning algorithms.	This study strengthens the extant literature by empirically analyzing restaurant reviews posted during the COVID-19 pandemic and discovering suitable deep learning algorithms for different text mining tasks.	Limited data availability, Causality limitations, Limited focus on qualitative factors

4	"COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients."[4]	Machine Learning Algorithms, including	The contribution of the work is the application of Big Data technologies and Machine Learning algorithms using Open Resource Libraries and Amazon Web Services (AWS) with the vision to improve the clinical diagnosis, even infectious disease with mathematical approaches.	Algorithm limitations, Limited data availability, Limited scope
5	"Teleworking—An Economic and Social Impact during COVID-19 Pandemic: A Data Mining Analysis".[5]	Data mining analysis	Identified based on job satisfaction and highlighted the most common employee profiles	Data limitations, Algorithm limitations, Limited scope
6	"Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes".[6]	Natural language processing techniques, Naïve Bayes model	The results yielded an 81.77% accuracy, which outweighs the accuracy of recent sentiment analysis studies using Twitter data from the Philippines. Accurate predictions of stock prices using machine learning algorithms.	The study's predictions and insights may be limited by the quality and quantity of data available.

7	“Small firms and the COVID-19 insolvency gap”.[7]	Statistical analysis.	In the ongoing crisis, policy makers should prefer instruments favoring entrepreneurs who respond innovatively to the pandemic instead of prolonging the survival of near-insolvent firms.	Geographical and data limitation.
8	"Studying the UK job market during the COVID-19 crisis with online job ads".	Statistical analysis.	This work presents an open methodology that enables a rapid and detailed survey of the job market in unsettled conditions and describes a web application jobtrender.com that allows others to query this data set.	The study's findings may not be generalizable to other countries or regions, as the impacts of COVID-19 on small businesses may vary depending on the local context.
9	"Working from home: small business performance and the COVID-19 pandemic."	Statistical analysis.	Find that in states with higher WFH rates, small businesses performed better overall with industry variations, controlling for the local pandemic, economic, demographic, and policy factors.	The limitation of this study is that it is focused solely.
10	"Predictive analytics using Big Data for the real	15 different machine	Out of the 15 different models	The research is based on

	estate market during the COVID-19 pandemic".	learning models	tested, extreme gradient boosting was the most accurate, although the difference was negligible	secondary data sources, which may have limitations in terms of accuracy and completeness.

From the above studies, we find that they have investigated various aspects related to the COVID-19 pandemic using different methodologies, such as machine learning algorithms, statistical analysis, and data mining techniques. Each study has highlighted its contributions, limitations, and findings. Some of the common limitations found in these studies are limited data availability, algorithm limitations, and limited scope. However, the studies have made valuable contributions to understanding consumer behavior, sentiment analysis, small business performance, and the real estate market during the pandemic. The findings of these studies have potential implications for policymakers, entrepreneurs, and individuals trying to navigate the ongoing crisis.

CHAPTER 3

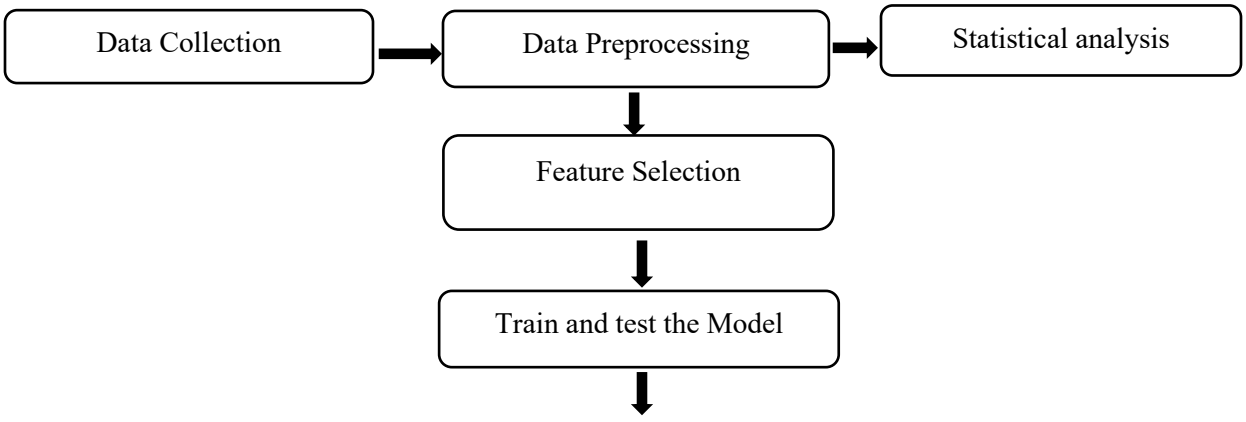
METHODOLOGY

3.1 INTRODUCTION

An overview of the research techniques used in the study is provided in this part. It contains details of the datasets and associated terms. The work is divided into two sections. The first section explains the process of statistically analyzing the datasets, and the second section explains the way to make predictions. The performance of the model is then examined using various metrics. We provide a description of the necessary tools for this research at the end of this chapter.

3.2 STRATEGY OF THE WORK

We have divided our workflow into a few sections. These include data collection, data preprocessing, statistical analysis, features selection for prediction, train and testing the model, and performance evaluation. We proceed through each module's subsequent part step by step, and those main modules are further split into a number of smaller sub-modules. Figure 3.1 illustrates the basic work stages.



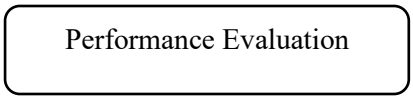


Figure 3.1: Workflow Diagram

3.3 DATA COLLECTION

We have gathered data from 150 business. This contains five distinct business categories. These include the bakery, poultry, meat, fruit, and sweet business. There are 30 businesses per category. First, we gathered the necessary data from markets located in Chatkhil and Noakhali Puoro Bazar in the Noakhali region. We conduct an offline survey to gather this data from the stores by handing out questionnaires. This survey takes into consideration about 18 parameters. In addition to business data, other personal information is factored in those parameters as well.

3.3.1 PARAMETERS

Table 3.1: Parameters

Variable name	**Value**
Name	Text
Business type	Text
Age	Number
Number of Employee before covid-19	Number
Number of Employee during covid-19	Number
Number of Employee after covid-19	Number
Business time before covid-19	Number
Business time during covid-19	Number
Business time after covid-19	Number
Sell before covid-19	Number
Sell during covid-19	Number
Sell after covid-19	Number
Expenditure before covid-19	Number
Expenditure during covid-19	Number
Expenditure after covid-19	Number
Profit before covid-19	Number
Profit during covid-19	Number
Profit after covid-19	Number

Here we consider 18 parameters from 150 business. We collected two personal information like name and age. The other parameters are described below:

(a) Business Type: Business Type refers what kind of product they can produce and sell. Here we conduct five types of businesses, such as, Food Bakery, Poultry business, Meat business, Fruit business, Sweet business. Food Bakery which produces Biscuits, Cake, Pau Ruti, Bread etc. The Poultry business is like Poultry Farm whose produce chicken and selling to the market. Meat business sell beep, mutton etc. The Fruit business which sells different types of sessional fruits. The Sweet business owner sell different types of sweet. We went all types of business and collected data.
(b) Number of Employee before covid-19: It refers that how many employees were kept for each type of business before the COVID-19 pandemic.
(c) Number of Employee during covid-19: It refers that how many employees were kept for each type of business during the COVID-19 pandemic.
(d) Number of Employee after covid-19: It refers that how many employees were kept for each type of business after the COVID-19 pandemic.
(e) Business time before covid-19: This parameter shows how long the business are opened before covid-19 pandemic.
(f) Business time during covid-19: This parameter shows how long the business are opened during covid-19 pandemic.
(g) Business time after covid-19: This parameter shows how long the business are opened after covid-19 pandemic.
(h) Sell before covid-19 : This is showing average product selling amount in a day before covid-19 pandemic.
(i) Sell during covid-19: This is showing average product selling amount in a day during covid-19 pandemic.
(j) Sell after covid-19 : This is showing average product selling amount in a day after covid-19 pandemic.
(k) Expenditure before covid-19: It is explained the expenditure to conduct a business per day before covid-19, which include product cost, employee salary, business rent, government tax etc.
(l) Expenditure during covid-19: It is explained the expenditure to conduct a business per day during covid-19, which include product cost, employee salary, business rent, government tax etc.
(m) Expenditure after covid-19 : It is explained the expenditure to conduct a business per day after covid-19, which include product cost, employee salary, business rent, government tax etc.
(n) Profit before covid-19: It describes the last amount which save after bearing all expenditure before covid-19 pandemic.
(o) Profit during covid-19: It describes the last amount which save after bearing all expenditure during covid-19 pandemic.
(p) Profit after covid-19: It describes the last amount which save after bearing all expenditure after covid-19 pandemic.

3.3.2 SAMPLE DATA

The following Table 3.2 shows the first nine rows of data that we collected from business. For properly utilizing the data, 17 features were selected in the final dataset. We will preprocess and train the models using this data. So, let's first take a closer look at our dataset.

Table 3.2: The first nine row of data

Business type	Bakery	Bakery	Bakery	Bakery	Bakery	Bakery	Bakery	The
Age	34	22	50	33	26	35	45	34
Number of Employee before covid-19	10	12	6	7	8	10	8	12
Number of Employee during covid-19	4	9	3	4	5	6	5	8
Number of Employee after covid-19	8	12	7	8	10	12	8	12
Business time before covid-19	10	12	10	9	8	12	10	10
Business time during covid-19	7	9	8	5	4	7	6	5
Business time after covid-19	12	12	10	10	10	12	10	12
Sell before covid-19	75000	80000	55000	65000	35000	70000	30000	50000
Sell during covid-19	55000	60000	30000	50000	25000	35000	15000	40000
Sell after covid-19	70000	85000	65000	80000	40000	10000	30000	70000
Expenditure before covid-19	57000	60000	37400	50050	26250	45000	24000	30000
Expenditure during covid-19	41800	45000	20400	40000	18750	24500	11000	25000
Expenditure after covid-19	53200	63750	42250	64000	29200	70000	22000	42000
Profit before covid-19	18000	20000	17600	14950	8750	25000	6000	20000
Profit during covid-19	13200	15000	9600	10000	6250	10500	4000	15000
Profit after covid-19	16800	21250	22750	16000	10800	30000	8000	28000

3.3.3 DATA PREPROCESSING

Data preparation is the process of preparing and cleaning raw data before analysis. The quality of the data will have an immediate influence on the accuracy of the results, making it a crucial stage in the data analysis process.

This is one of the most important stages of the implementation module. We could get a lot of information from many sources, therefore some of it might be inaccurate or null. There may on sometimes be values in the dataset that are unclear, redundant, or missing.

Additionally, some data may have a lot of dimensions, which is problematic since it makes training much harder. Because of this, data preparation is crucial to this process. The processing of duplicates, missing values, dimension reduction, etc. is covered in data preparation methods. The following Figure displayed below indicates the steps involved in data preprocessing.

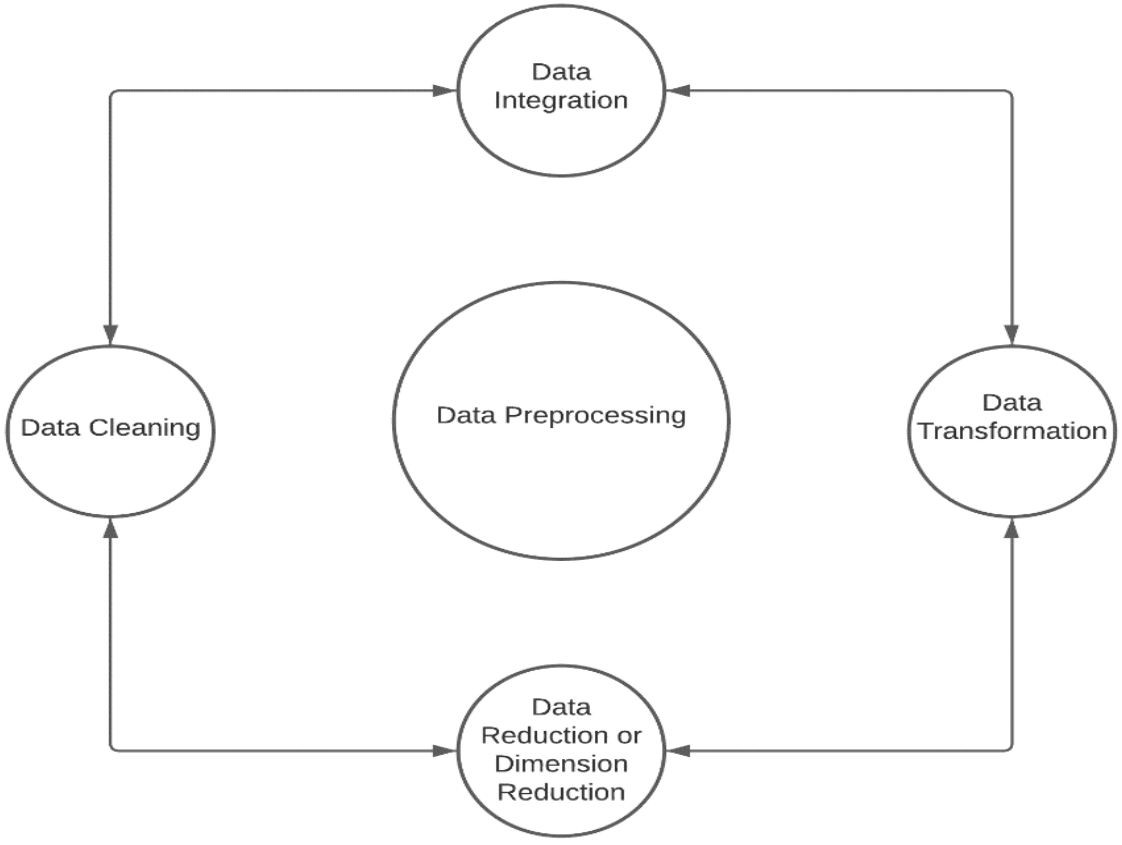


Fig 3.2: Tasks Involved in Data Preprocessing.

a) **Data Cleaning:** Both the Pandas and Scikit-Learn packages make it simple to implement the preprocessing of the dataset. Methods in the Pandas library, such as rename(), replace(), dropna(), loc(), iloc(), and drop(), can be used for preprocessing and Strong classes in Scikit-Learn, such as PCA and LDA, may effectively minimize the dimension of the dataset. The dropna() function was used to deal with duplicate or missing variables in our collection. Due to the small size of the dataset dimension, training time will not be much impacted. Therefore, dimensional reduction is not necessary. However, we will change a few category aspects. To employ numerical features, use the replace() function to replace values with the average of the ranges mentioned in the survey question. Unfortunately, the dataset contains certain values that are not entirely obvious. In order to avoid confusing the model while it is being trained, we have substituted such numbers with values that are the closest to those in our surveys. However, doing so would lower the already sparse amount of data we have. There were no issues with the dataset other from that. We can now proceed to the following action.

b) **Feature scaling:** Normalizing the dataset's input variables entails feature scaling. Scaling the features of the data is one of the most crucial changes that must be applied. Machine Learning algorithms typically struggle to perform effectively when the scales of the numerical attributes are extremely varied, with very few exceptions. Hence, scaling the data becomes essential. It is possible to address this issue by standardization. With the help of this technique, some data with a scale of 1 to 1000 can be converted into a value between 0 and 1. The machine learning algorithms' capacity to anticipate outcomes is enhanced. It's actually very easy to standardize something. To ensure that standardized values always have a zero mean, it first subtracts the attribute's mean value from each of the attribute's values. Next, it divides by the variance to create a distribution with a unit variance. This is the best method since standardization is considerably more resistant to the effects of outliers.

$$X_{Normalized} = x\text{-}min(x)/max(x)\text{-}min(x) \tag{3.1}$$

Python's Scikit-Learn library has classes named StandardScaler, MinMaxScaler, and RobustScaler which perform standardization.

c) **Splitting the Dataset:** Every dataset cannot be used at once. The dataset is frequently split into two sets. Data sets are split into training and testing sets as part of the data assessment process mining modeling, A dataset is frequently split into a training set and testing set, with most of the data being utilized for training and the remaining portion being kept for later testing. In order to assist ensure that the training and testing sets have an adequate range of values, this service's tools randomly sample the data. By using comparable data for training and testing, the effects of data inconsistencies may be minimized, and the model's characteristics can be better understood. The training set will contain most of the dataset. The models will be trained using 70% of the dataset. For testing purposes, the remaining 30% will be maintained in stock. After being processed using the training set, a model is evaluated by making predictions against the test set. Because the data in the testing set already has known values for the feature that needs to be predicted, it is easy to determine if the model's predictions are true. Testing is the process's last step. On the testing set, only the trained

models' capacity to generalize to the dataset is examined. To divide the dataset, we utilized the Scikit-Learns train test split class.

3.4 STATISTICAL ANALYSIS

A sub-field of mathematics called statistical analysis is concerned with the gathering, interpretation, analysis, and presentation of data. Making inferences and reaching conclusions about a population based on a sample of data is the basic goal of statistical analysis. To evaluate data and derive meaningful conclusions, statistical analysis employs statistical techniques including descriptive statistics, inferential statistics, hypothesis testing, regression analysis, and others. Decision-making in a variety of sectors, including business, medical, economics, social sciences, and more, may be supported by statistical analysis results.

Here, we have 5 types of business data. We make analysis in those business. We analyze the data in two ways. Such as, Individual Business and Considering all Business

(a)Individual Business: We analyze the Bakery business, Poultry business, Meat business, Fruit business, and Sweets business data individually. In each business, we collect same parameters for all types of business like as, Number of employee, Business time, Product selling amount, Business expenditure, Profit in time of before, during, after COVID-19. We visualize the total number of employee before, during, and after COVID-19 pandemic in a bar chart with mathplot.pyplot(), bar() plot function. We also visualize the mean value of the Business time, Product selling amount, Business expenditure, Profit. Next, we visualize the Selling amount, Expenditure amount, Profit with the line plot() function. From the visual representation of the data, we can see, how COVID-19 impacts the local business performance.

(b) Considering all Business: Next, we visualize the Bakery business, Poultry business, Meat business, Fruit business, and sweets business data in combine. We visualize each parameter for all types of business in each bar graph. Such as, first we visualize the total Number of employees before, during, and after COVID-19 for different business in a bar graph. Similarly, we also visualize Business time, Product selling amount, Business expenditure, Profit in time of before, during, after COVID-19. We also plot the Selling amount, Expenditure and Profit for 150 business with line plot() function. From all visualization, we can see how the COVID-19 impacts on these 5 types of business, and which types of business are more affected and which is less affected.

For these statistical analysis we use Python. Which has many libraries that can be used for analysis, including NumPy, Pandas, SciPy, Statsmodels, and Scikit-learn.

3.5 PREDICTION

3.5.1 PROFIT PREDICTION AND FEATURES SELECTION

Profit prediction is the technique of estimating future profitability of a company or organization using statistical or machine learning models. In order to estimate future income, costs, and profits, this procedure entails evaluating previous financial data as well as other pertinent data to find patterns and trends. Regression analysis, time series analysis, and machine learning algorithms like neural networks and decision trees are just a few statistical methods that may be used to anticipate profits. These methods can be used to pinpoint important variables that are likely to affect profitability, such as alterations in the market environment, modifications in consumer behavior, or adjustments in manufacturing costs. It is crucial to employ high-quality data and to routinely update models based on new information in order to create precise profit estimates. Profit forecasting is a useful tool that may be used by companies and organizations across a wide range of sectors to guide decision-making and simplify long-term planning. In this section, we predict the profit for businesses. We predict three categories profit individually for the purpose of COVID-19. These three categories are before, during, and after COVID-19 Profit. To predict this Profit, we use Machine Learning algorithm. From different types of ML algorithms, we choose three types of regression algorithms. These are Linear Regression algorithm, Random Forest Regression, K-Nearest Neighbor. From these algorithms, we find out how much these models are fit with our dataset. We also measured these models' performance with, performance metrics. Now, for predicting Profit we need to select the best features. In the next section we describe Feature Selection process, Machine Learning Algorithms, and the implementation process of these algorithms.

In the real world, it is rare to build a machine learning model using all the variables in a dataset. The inclusion of useless or redundant information affects the model's ability to generalize, which may also reduce a classifier's overall accuracy. As additional variables are included in a model, its total complexity rises. The least number of assumptions is essential for the optimal solution to a problem. As a result, choosing features for machine learning models is an important step. The goal of feature selection in machine learning is to identify the best collection of characteristics for creating useful models of the phenomena being researched. To identify traits that will improve the efficiency of supervised learning, labeled data can be subjected to machine learning feature selection approaches. Here we predict Before COVID-19 profit, During COVID-19 profit, and After COVID-19 profit. Now we can select features for three types.

For Before COVID-19 profit prediction**:** We select 4 features. These are-

1. Number of Employee (Before covid-19)
2. Total business time (Before covid-19)
3. Sell Before covid-19
4. Expenditure Before covid-19

For During COVID-19 profit prediction**:** We select 4 features. These are-

1. Number of Employee (During covid-19)
2. Total business time (During covid-19)
3. Sell During covid-19
4. Expenditure During covid-19

For After COVID-19 profit prediction: We select 4 features. These are-

1. Number of Employee (After covid-19)
2. Total business time (After covid-19)
3. Sell After covid-19
4. Expenditure After covid-19

3.5.2 PREDICTION ALGORITHMS

In this Section, we briefly describe about the algorithms. Here we used Supervised learning algorithms. Supervised learning is a type of machine learning in which a model is trained using labeled data, meaning that each sample in the training set is associated with a known output or target value. Supervised learning algorithms aim to learn the relationship between the input data (features) and the output data (labels or target values) by finding patterns or rules that enable accurate predictions on new, unseen data. There are several types of supervised learning algorithms,

Here we use regression algorithms because we have numerical data, such as Selling amount, Expenditure, and profit. We use three regression algorithms, such as Linear Regression, Random Forest regression, and K-Nearest Neighbor. The next subsections will cover all these algorithms.

(i) Linear Regression

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. The linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:

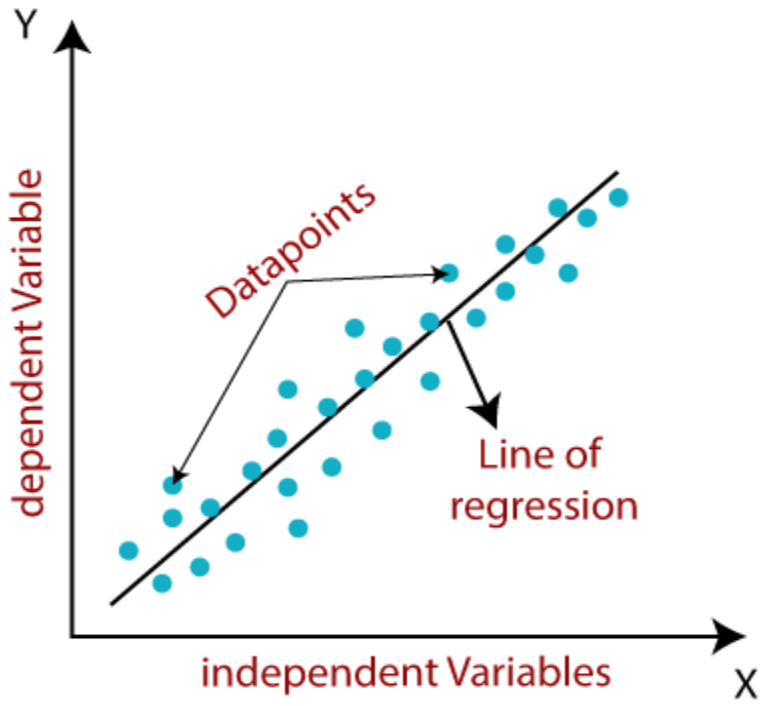


Figure 3.3: Linear Regression

Mathematically, we can represent a linear regression as:

$y= b0+b1x+ \varepsilon$ ----------(2)

Here,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

b0= intercept of the line (Gives an additional degree of freedom)

b1 = Linear regression coefficient (scale factor to each input value).

ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

For our dataset we can make a equation for this algorithm. Here "Profit" is the dependent variable, and "Number of Employee", "Business Time", "Sell", and "Expenditure" are the independent variables, then the equation for multiple linear regression can be written as:

$Profit = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta3X_3 + \beta4X_4 + \varepsilon$

----------(3)

Here , X_1 = Number of Employee

X_2 = Business Time

X_3 = Sell

X_4 = Expenditure

Also, β0 is the intercept, β1, β2, β3, and β4 are the slopes, and ε is the error term. The intercept represents the expected value of the dependent variable when all independent variables are equal to zero. The slopes represent the change in the dependent variable associated with a one-unit increase in each independent variable, holding all other variables constant.

(ii) Random Forest

Random Forest is a supervised learning algorithm. It creates a "forest" out of an ensemble of decision trees, which are commonly trained using the "bagging" method. The bagging method's basic premise is that combining several learning models improves the overall output. The Random Forest algorithm introduces extra randomness when splitting a node, it searches for the best feature among a random subset of features, this results in a greater tree diversity, which trades a higher bias for a lower variance, generally yielding an overall better model. So, the prerequisites for the random forest to perform well are: There needs to be some actual signal in our features so that models built using those features do better than random guessing. Also, the predictions (and therefore the errors) made by the individual trees need to have low correlations with each other. Some of the features of this algorithm are stated below:

1. It takes less training time as compared to other algorithms.

2. It predicts output with excellent accuracy, and it runs quickly even with a huge dataset.

3. It can also maintain accuracy if a significant amount of data is absent.

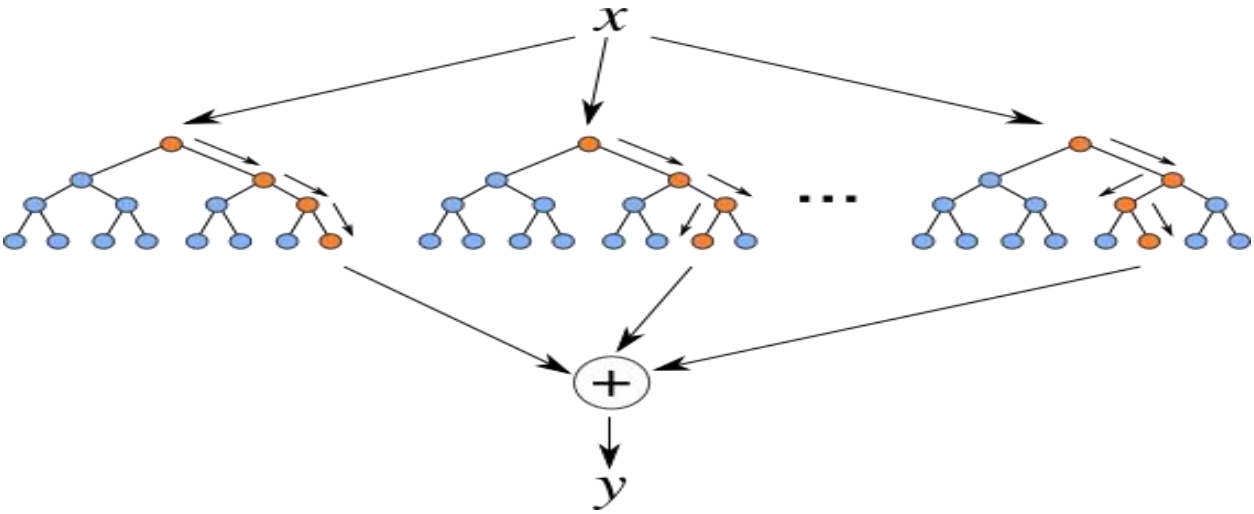


Figure 3.4: Random Forest

The random forest is formed in two phases: the first is to combine the N decision trees to build the random forest, and the second is to make predictions for each tree created in the first phase. The following steps can be used to demonstrate the working process:

Step - 1. Pick K data points at random from the training set.

Step - 2. Create decision trees for the data points you've chosen (Subsets).

Step - 3. Decide on the number N for the decision trees you wish to create.

Step - 4. Repetition of Steps 1 and 2.

Step - 5. Find the forecasts of each decision tree for new data points and allocate the new data points to the category with the most votes.

(iii) K-Nearest Neighbor

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. KNN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most like the available categories. KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using KNN algorithm. KNN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. KNN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

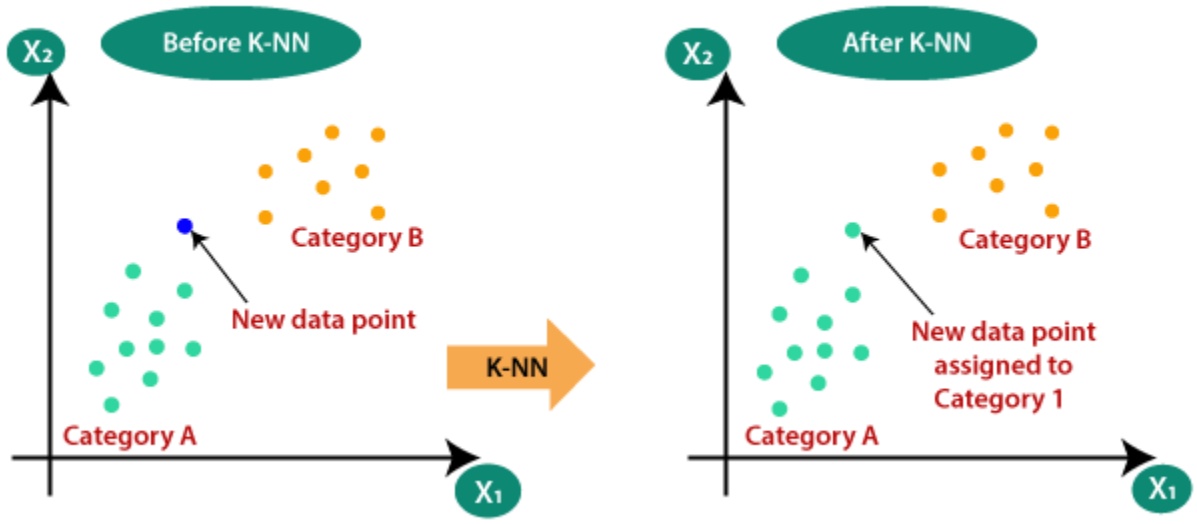


Figure 3.5: K-Nearest Neighbor

The KNN working can be explained based on the below algorithm:

Step-1: Select the number K of the neighbors.

Step-2: Calculate the Euclidean distance of K number of neighbors.

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.

Step-6: Our model is ready.

3.6 PERFORMANCE ANALYSIS

The performance analysis involves systematic observations to enhance performance that improves the decision-making of data and gives feedback to our system. It is the process of studying or evaluating the performance of a particular scenario in comparison to the objective to be achieved. For analysis, we use R2 score, mean_absolute_score, mean_squared_error.

(i) R2 score

R-squared is a statistical method that determines the goodness of fit. It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%. The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model. It is also called a coefficient of determination, or coefficient of multiple determination for multiple regression.

It can be calculated from the below formula:

$$R-Squared=\frac{Explained\,\text{var}iation}{Total\,\text{var}iation} \quad \text{----------(4)}$$

(ii) Mean Absolute Error (MAE)

Mean Absolute Error or MAE is one of the simplest metrics, which measures the absolute difference between actual and predicted values, where absolute means taking a number as Positive. To understand MAE, let's take an example of Linear Regression, where the model draws a best fit

line between dependent and independent variables. To measure the MAE or error in prediction, we need to calculate the difference between actual values and predicted values. But in order to find the absolute error for the complete dataset, we need to find the mean absolute of the complete dataset.

The below formula is used to calculate MAE:

$$MAE = \frac{1}{N}\sum |Y - Y'| \text{----------(5)}$$

Here,

Y is the Actual outcome, Y' is the predicted outcome, and N is the total number of data points.

MAE is much more robust for the outliers. One of the limitations of MAE is that it is not differentiable, so for this, we need to apply different optimizes such as Gradient Descent. However, to overcome this limitation, another metric can be used, which is Mean Squared Error or MSE.

(iii) Mean Squared Error

Mean Squared error or MSE is one of the most suitable metrics for Regression evaluation. It measures the average of the Squared difference between predicted values and the actual value given by the model. Since in MSE, errors are squared, therefore it only assumes non-negative values, and it is usually positive and non-zero. Moreover, due to squared differences, it penalizes small errors also, and hence it leads to over-estimation of how bad the model is. MSE is a much-preferred metric compared to other regression metrics as it is differentiable and hence optimized better.

The formula for calculating MSE is given below:

$$MSE = \frac{1}{N}\Sigma |Y - Y'|^2 \text{----------(6)}$$

Here, Y is the Actual outcome, Y' is the predicted outcome, and N is the total number of data points.

3.7 REQUIRED TOOLS

Different tools will be used for this study. All of them are free and open source. A short description of these tools is given below,

a) Python: Python is a high-level general programming language.
b) Virtual Environment: We will use Jupyter Notebook as the environment to conduct our experiment. Anaconda framework provides Jupyter Notebook along with other tools that

help in conducting such experiments. We will create a virtual environment inside Anaconda and install all the required tools.

c) Scikit-learn: Scikit-learn is a widely used popular library for machine learning.
d) NumPy: NumPy is a very powerful package that enables us for scientific computing.
e) Matplotlib: Matplotlib is a plotting library for Python and its numerical mathematics extension NumPy.
f) Pandas: Pandas is an open-source BSD licensed software specially written for python.
g) Seaborn: Seaborn is a library used for data visualization and is created by using python.
h) SciPy: SciPy is a collection of fundamental mathematical functions and is built on the top of Scikit-learn.
i) OS: The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system dependent functionality.
j) Skopt: Scikit-Optimize, or skopt for short, is an open-source Python library for performing optimization tasks. It offers efficient optimization algorithms, such as Bayesian Optimization, and can be used to find the minimum or maximum of arbitrary cost functions.

CHAPTER 4

RESULTS AND ANALYSIS

4.1 INTRODUCTION

By examining the results of the statistical analysis using bar charts, correlation matrices, and other visual tools, we evaluate the dataset in this chapter. The part that follows discusses this. In the following part, the results of the numerical analysis using prediction algorithms are discussed. We will draw a conclusion after evaluating the results of several performance measures. The regression model with the greatest generalization performance for the dataset will be identified by comparing these values. In two different methods, we analyze the datasets through visualization. The information is individually and categorically analyzed in the first section. The second part combines the five datasets to begin the analysis.

4.2 STATISTICAL ANALYSIS FOR IMPACTS OF COVID-19

The COVID-19 pandemic has had a significant effect on many industries, causing job losses, financial losses, and changes in customer behavior. Businesses will need to adapt and innovate if they want to thrive in the post-COVID-19 pandemic. The impacts of COVID-19 will probably last for years to come.

In this section, statistical analysis is done on the information. Statistical analysis is an important process that involves performing preliminary analyses on data in order to discover patterns, spot anomalies, test theories, and confirm presumptions using summary statistics and graphical representations. It is advisable to fully understand the material before attempting to draw as many conclusions as we can from it.

4.2.1 IMPACTS ANALYSIS FOR INDIVIDUAL BUSINESS

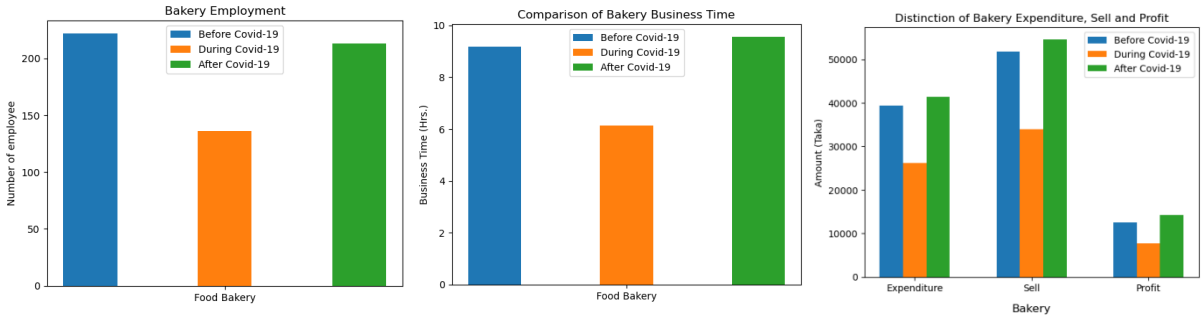
(a)Bakery Business

To perform a descriptive analysis of this business, we examine the different variables and their distributions, measures of central tendency, and dispersion. First, let's look at the continuous variables. These are Age, Number of Employee (Before, During, and After Covid-19), Total business time (Before, During, and After Covid-19), Sell (Before, During, and After Covid-19), Cost (Before, During, and After Covid-19), Profit (Before, During, and After Covid-19). We have a total of 30 observations for Age with a minimum age of 22, a maximum age of 50, and an average age of approximately 32.57 years.□ The average number of employees for all three phases is approximately 5.57, 3.95, and 6.24, respectively. The minimum number of employees for each phase is 2, and the maximum number is 12.□The average number of business times for each phase is approximately 9.38, 6.05, and 10.10, respectively. The minimum number of business times for each phase is 4, and the maximum is 12.The average daily sell for each phase is approximately 49,866.67, 34,166.67, and 53,166.67, respectively. The minimum daily sell for each phase is 15,000, and the maximum daily sell is 100,000.The average daily expenditure for each phase is

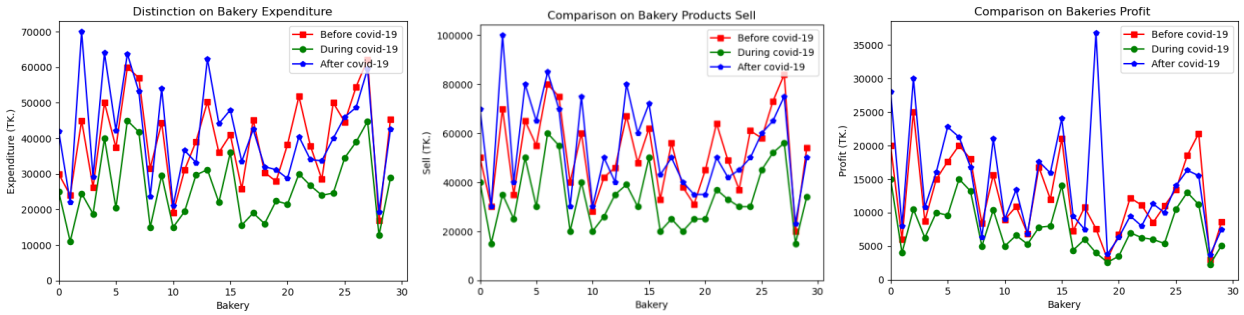
approximately 24,192.86, 19,807.14, and 27,773.81, respectively. The minimum daily expenditure for each phase is 11,000, and the maximum daily expenditure is 45,000.The average daily profit for each phase is approximately 25,673.81, 14,359.52, and 25,392.86, respectively. The minimum daily profit for each phase is 3,500, and the maximum daily profit is 35000.

Based on the statistical analysis, it appears that the business experienced a significant drop in sales and profits during the pandemic, but has since rebounded and even surpassed pre-pandemic levels. The mean and median sales during pandemic were lower than before the pandemic, with a percentage difference of -34.64%. However, after the pandemic, sales increased and had a positive percentage difference of 5.39%. The paired t-test for sales during pandemic showed a significant difference from before the pandemic (T-Statistic = 12.45, P-Value = 3.71e-13), indicating that the drop in sales during the pandemic was not due to chance. However, the paired t-test for sales after the pandemic did not show a significant difference from before (T-Statistic = -1.47, P-Value = 0.153), suggesting that the increase in sales after the pandemic may not be statistically significant. Similarly, the business experienced a significant drop in expenditure during pandemic, with a percentage difference of -33.64%. However, after the pandemic, expenditure increased and had a positive percentage difference of 5.07%. The paired t-test for expenditure during pandemic showed a significant difference from before the pandemic (T-Statistic = 11.90, P-Value = 1.11e-12), indicating that the decrease in expenditure during the pandemic was not due to chance. The paired t-test for expenditure after the pandemic did not show a significant difference from before (T-Statistic = -1.35, P-Value = 0.186), suggesting that the increase in expenditure after the pandemic may not be statistically significant. Finally, the business experienced a significant drop in profits during pandemic, with a percentage difference of -37.79%. However, after the pandemic, profits increased and had a positive percentage difference of 14.12%. The paired t-test for profits during pandemic showed a significant difference from before the pandemic (T-Statistic = 8.95, P-Value = 7.74e-10), indicating that the drop in profits during the pandemic was not due to chance. The paired t-test for profits after the pandemic did not show a significant difference from before (T-Statistic = -1.61, P-Value = 0.118), suggesting that the increase in profits after the pandemic may not be statistically significant.

Overall, the statistical analysis suggests that the business was negatively affected by the pandemic, but has since rebounded and even surpassed pre-pandemic levels. However, it is important to note that the increase in sales, expenditure, and profits after the pandemic may not be statistically significant. It may be worthwhile for the business to continue monitoring these metrics and conducting further analysis to identify areas for improvement.



(a)



(b)

Figure 4.1: (a) Bakery Business Employment, Business Time, Expenditure, Sale and Profit

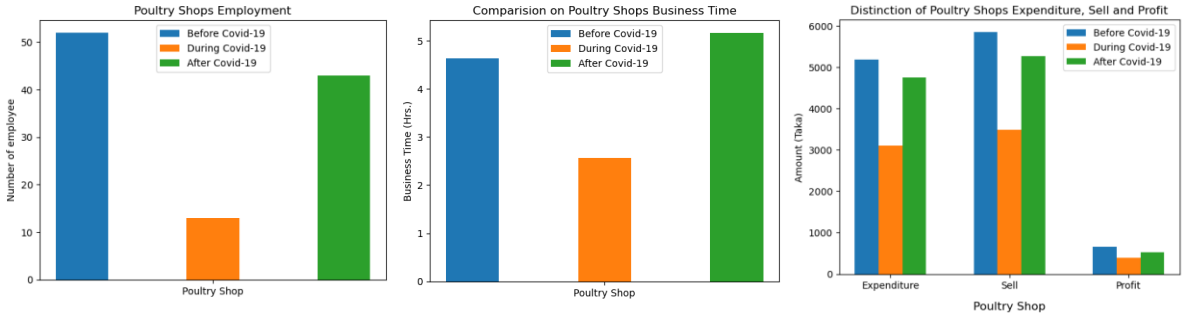
(b)Poultry Business

The descriptive analysis for Poultry business are given.The data set consists of 30 observations of a Poultry business, which provides information about various parameters related to its business, such as age of the business, number of employees, total business times, sales, cost, and profit, before, during, and after the pandemic. The mean age of the businesses is 32 years, ranging from 25 to 45 years. The average number of employees before, during, and after the pandemic are 1.4, 0.4, and 1.3, respectively. This indicates that most businesses had to lay off some of their employees during the pandemic. The average total business times before, during, and after the pandemic are 4.3, 2.5, and 5.1, respectively. The increase in business times after the pandemic could be due to businesses trying to recover the losses incurred during the pandemic. The average sales before, during, and after the pandemic are 5848, 3531, and 5491, respectively. This indicates a decrease in sales during the pandemic, but a subsequent recovery after the pandemic. The average cost before, during, and after the pandemic are 4811, 3298, and 4796, respectively. The decrease in cost during the pandemic could be due to a decrease in demand, but the cost has returned to normal after the pandemic. The average profit before, during, and after the pandemic are 1037, 233, and 695, respectively. The sharp decline in profit during the pandemic indicates that most businesses were hit hard by the pandemic, but there was a recovery in profit after the pandemic.

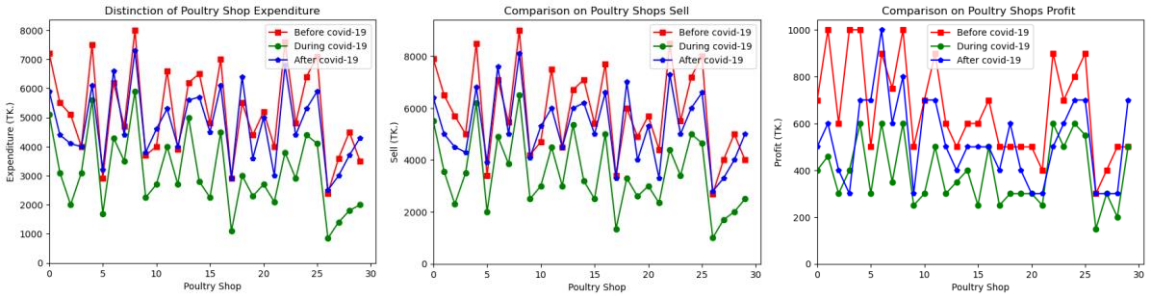
Based on the statistical analysis provided, it appears that the business experienced a significant decrease in sales, expenditure, and profits during the pandemic compared to before and after the pandemic. The mean and median values for sales, expenditure, and profits were all higher before the pandemic and gradually increased after the pandemic. However, during the pandemic, there was a significant decline in these values. The percentage difference for sales, expenditure, and profits during the pandemic was -40.44%, -40.30%, and -41.55%, respectively. These figures suggest that the pandemic had a substantial negative impact on the business's financial performance. It is important to note that the paired t-tests for sales, expenditure, and profits during and after the pandemic resulted in NaN values. This may be due to insufficient data or other factors

that require further investigation. Therefore, it is difficult to determine whether the observed differences in sales, expenditure, and profits are statistically significant.

In conclusion, the statistical analysis suggests that the pandemic had a significant negative impact on the business's financial performance. The business may need to adapt to the changing economic environment and implement strategies to mitigate the impact of future crises. It is also important to conduct further analysis and gather additional data to fully understand the extent of the pandemic's impact on the business.



(a)



(b)

Figure 4.2: Poultry Business Employment, Business Time, Expenditure, Sale and Profit

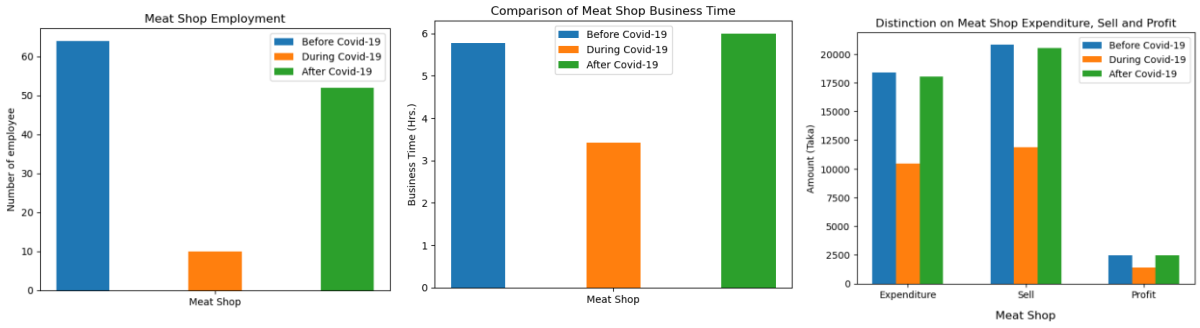
(c)Meat Business

This graph provides a detailed analysis of meat business in terms of various business metrics before, during, and after the pandemic. The age range of employees in the business ranges from 22 to 50 years old, with an average age of 32.48 years. Before the pandemic, the minimum number of employees was one, while the maximum was four, with an average of 2.04 employees. During the pandemic, the average number of employees dropped significantly, with a minimum of zero

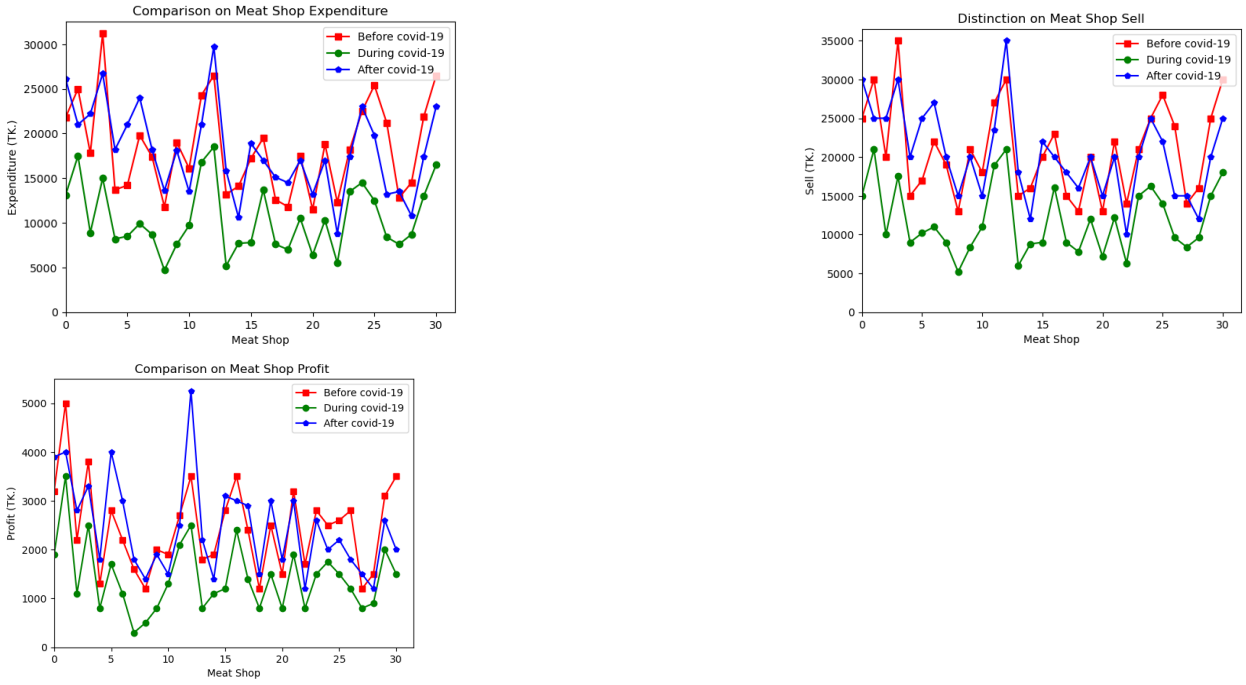
and a maximum of one, with an average of 0.23 employees. However, after the pandemic, the average number of employees increased to 1.56, with a minimum of one and a maximum of three. Before the pandemic, the business was operational for an average of 6.04 hours per day, with a range of four to eight hours. During the pandemic, the average operational hours dropped to 3.27 hours, with a range of three to four hours. After the pandemic, the business resumed operating for an average of 5.31 hours per day, with a range of five to eight hours. In terms of sales per day, the business recorded an average of 21,013.64 before the pandemic, with a range of 13,000 to 35,000. During the pandemic, the average sales per day dropped to 10,234.09 , with a range of 5,200 to 21,000 . After the pandemic, the business recorded an average of 20,277.27 in sales per day, with a range of 10,000 to 30,000. Before the pandemic, the business spent an average of 18,394.09 per day on expenditure, with a range of 11,500 to 31,200 . During the pandemic, the average expenditure per day dropped to 9,027.73 , with a range of 4,700 to 17,500 . After the pandemic, the business spent an average of 17,725 per day on expenditure, with a range of 13,600 to 26,700. In terms of profit per day, the business recorded an average of 2,624.55 before the pandemic, with a range of 1,200 to 3,800 . During the pandemic, the average profit per day dropped to 1,206.36 , with a range of 300 to 2,500 . After the pandemic, the business recorded an average profit per day of 2,552.27, with a range of 1,400 to 5,250 . These metrics provide valuable insight into the business's performance and how it was affected by the pandemic.

Based on the statistical analysis , it appears that the business experienced a significant decrease in sales, expenditure, and profit during the pandemic. The mean sales before the pandemic was 20,838.71 per day, which dropped to 11,853.23 during the pandemic, a decrease of 43.12%. Similarly, the expenditure before the pandemic was 18,390.32 per day, which decreased to 10,435.48 per day during the pandemic, a decrease of 43.26%. However, after the pandemic, both sales and expenditure increased, and they were close to their pre-pandemic levels. The mean sales after the pandemic was 20,500 per day, which was only 1.63% lower than the pre-pandemic level. Similarly, the mean expenditure after the pandemic was 18,043.55 per day, which was only 1.89% lower than the pre-pandemic level. The paired t-tests conducted for sales, expenditure, and profit show that the differences between the means during and after the pandemic are not statistically significant. However, the t-tests for sales, expenditure, and profit during the pandemic show a statistically significant difference between the means before and during the pandemic.

Overall, it appears that the business was significantly affected by the pandemic, experiencing a drop in sales, expenditure, and profit. However, it was able to recover somewhat after the pandemic. The paired t-tests suggest that the recovery was not statistically significant, but it is still encouraging to see that the business was able to bounce back somewhat after the pandemic.



(a)



(b)

Figure 4.3: Meat Business Employment, Business Time, Expenditure, Sale and Profit

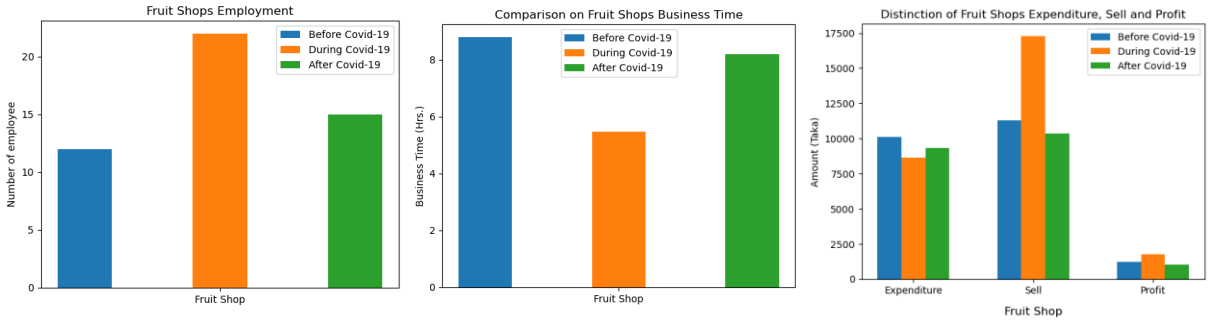
(d)Fruit Business

The graphs contains data related to various aspects of a Fruit business, including the age of the business, the number of employees before, during and after pandemic, the total business times before, during and after Covid-19, the daily sales, costs and profits before, during and after pandemic, and the type of business. The minimum age of the business is 23 years, while the maximum age is 45 years. The number of employees before, during and after pandemic ranges

from 0 to 1 for each category. The total business times before, during and after pandemic range from 6 to 12 times per day for each category. The daily sales before, during and after pandemic range from 6,000 to 17,000, 10,000 to 30,000 and 7,000 to 20,000, respectively. The daily costs before, during and after pandemic range from 5,100 to 15,300, 4,500 to 18,000 and 6,300 to 13,100, respectively. The daily profits before, during and after pandemic range from 700 to 1,600, 700 to 3,000 and 700 to 2,000, respectively.

Based on the analysis, it seems that the company experienced a significant increase in sales during the pandemic period compared to before, with a percentage difference of 53.1%. However, after the pandemic, the sales decreased by 8.3% compared to the pre-pandemic period. The paired t-test showed that there was a significant difference in sales during pandemic, but not after. In terms of expenditure, there was a decrease in costs during pandemic, with a percentage difference of 14.7%. However, after the pandemic, the costs remained lower but only decreased by 7.7% compared to pre-pandemic levels. The paired t-test showed a significant difference in expenditure during pandemic, but not after. The company also experienced an increase in profits during pandemic, with a percentage difference of 46.6%. However, after the pandemic, the profits decreased by 12.9% compared to pre-pandemic levels. The paired t-test showed a significant difference in profits during pandemic, but not after.

Overall, it seems that the pandemic had a significant impact on the company's sales, expenditure, and profits, but these effects were not sustained after the pandemic. The company may need to adjust its strategies to maintain the increased sales and profits seen during COVID-19.



(a)

Figure 4.4: Fruit Business Employment, Business Time, Expenditure, Sale and Profit

(b)

Figure 4.4: Fruit Business Employment, Business Time, Expenditure, Sale and Profit

(e)Sweet Business

The graph provided valuable insights into the performance of a Fruit business during the three periods - before, during, and after Covid-19. The sales and profit figures for the business remained consistent over the three periods, with higher sales and profits during pandemic, while the expenditure was higher during the same period. The number of employees and business times remained relatively stable, with little variance across the three periods. An analysis of the minimum, maximum, and mean values reveals that the mean age of employees is 32.6 years, with the youngest employee being 23 years old and the oldest being 45 years old. The mean number of employees before pandemic is 3.2, with a maximum of 5 and a minimum of 0. During pandemic, the mean number of employees is 1.3, with a maximum of 2 and a minimum of 0, while after pandemic, the mean number of employees is 4.8, with a maximum of 10 and a minimum of 4. The total business times showed little variation across the periods, with a mean of 11.8 times before, 4.8 times during, and 11.1 times after the pandemic. The sales figures show that the mean sales amount before is 7,975, with a minimum of 5,000 and a maximum of 12,000. During pandemic, the mean sales amount is 4,415, with a minimum of 2,000 and a maximum of 7,000, while after pandemic, the mean sales amount is 7,450, with a minimum of 5,000 and a maximum of 10,000. The costs incurred by the business show little variance over the periods, with a mean cost of 3,075 before, 2,250 during, and 1,425 after the pandemic. Finally, the mean profit before pandemic is 4,900, with a minimum of 1,000 and a maximum of 3,400. During pandemic, the mean profit amount is 2,165, with a minimum of 450 and a maximum of 1,050, while after pandemic, the mean profit amount is 6,025, with a minimum of 400 and a maximum of 1,800. Overall, the graph shows that the Fruit business maintained stable employee numbers and business times over the three periods, while experiencing fluctuations in sales, expenditure, and profit figures.

Based on the information provided, we can make several observations. The mean and median sales during and after pandemic are lower than before the pandemic, and there is a statistically significant difference in sales during and after pandemic compared to before the pandemic based on paired t-tests. The percentage difference in sales during and after pandemic compared to before the pandemic is -45.75% and -8.10% respectively. The mean and median expenditure during and after pandemic are lower than before the pandemic, and there is a statistically significant difference in expenditure during and after pandemic compared to before the pandemic based on paired t-tests. The percentage difference in expenditure during and after pandemic compared to before the pandemic is -42.66% and -22.98% respectively. The mean and median profit during and after pandemic are lower than before the pandemic, and there is a statistically significant difference in profit during pandemic compared to before the pandemic based on a paired t-test. The percentage

difference in profit during and after pandemic compared to before the pandemic is -58.10% and -8.10% respectively.

Overall, the graph shows that the business experienced a decline in sales, expenditure, and profit during and after the pandemic compared to before the pandemic. The difference in sales and expenditure during and after COVID-19 compared to before the pandemic is less severe than the difference in profit, which may indicate that the business had difficulty maintaining profitability during the pandemic.

(a)

(b)

Figure 4.5: Sweet Business Employment, Business Time, Expenditure, Sale and Profit

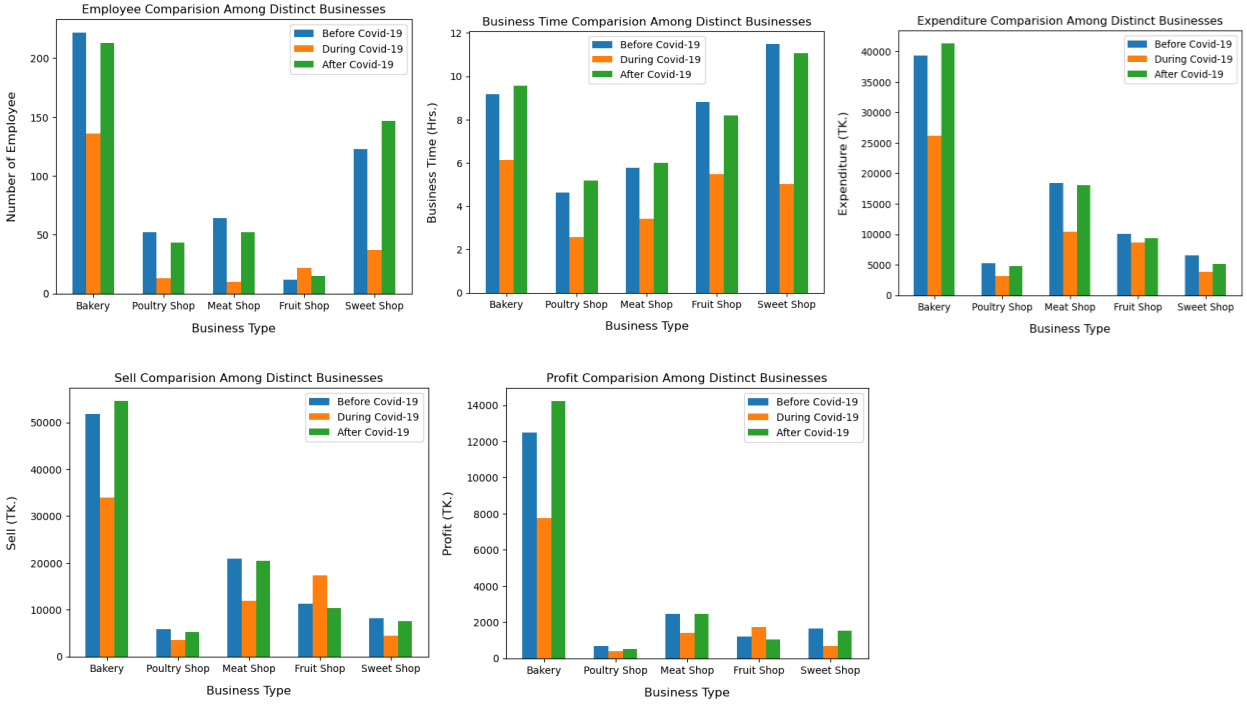
4.2.2 IMPACTS ANALYSIS FOR ALL BUSINESS

Comparing all the businesses based on the information, we can see that the pandemic had a varying impact on each business. The fruit business had the most significant positive growth trend during the pandemic, with a percentage difference of 53.10%, indicating that it was likely in high demand during this time. However, this growth trend declined after the pandemic, with a percentage difference of -8.26%. The bakery business had a significant decrease in mean sell during the pandemic, with a percentage difference of -34.64%, but showed a positive growth trend after the pandemic, with a percentage difference of 5.39%. Both the poultry and meat businesses experienced a considerable decrease in mean sell during the pandemic, with percentage differences

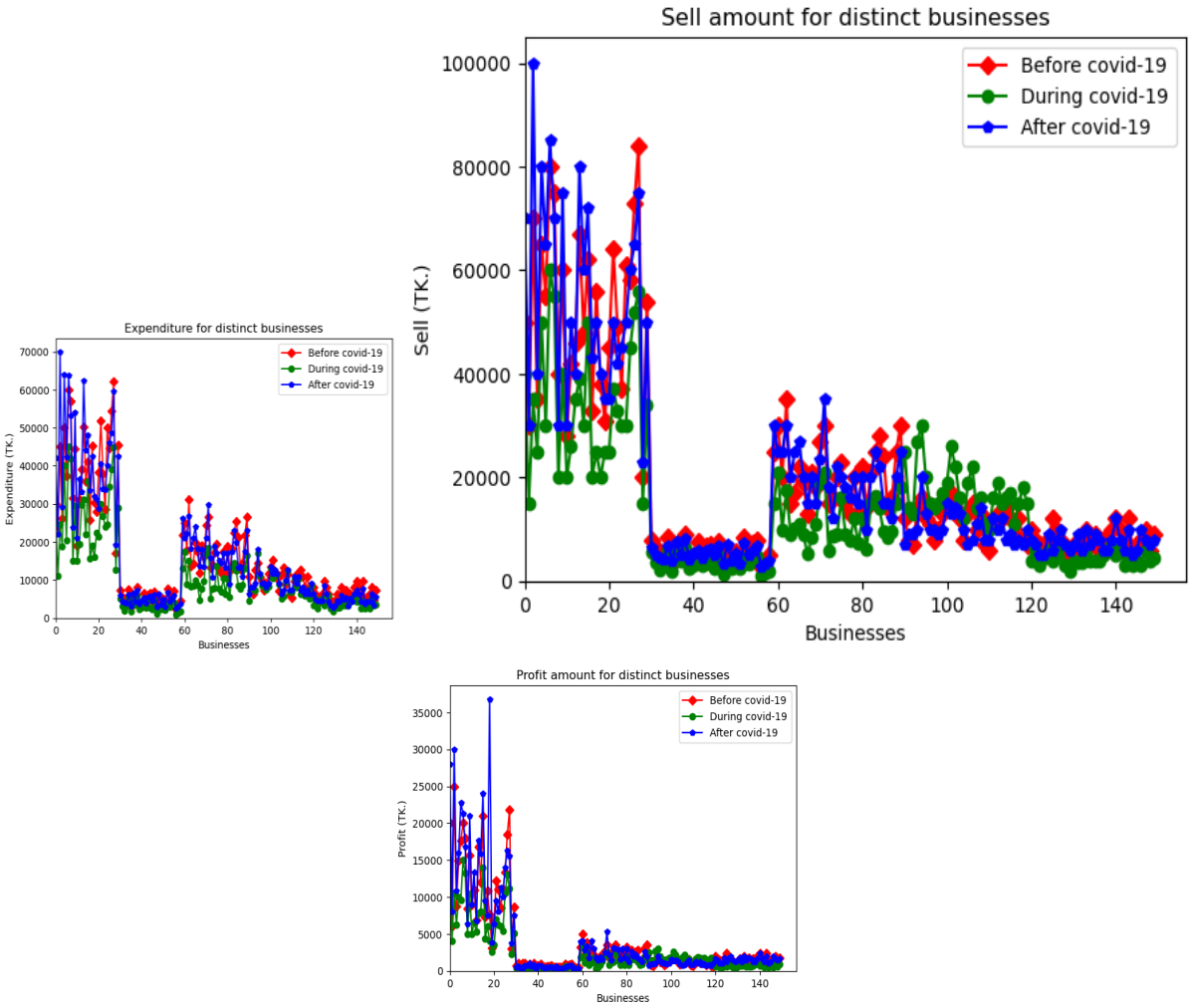
of -40.44% and -43.12% respectively. After the pandemic, the percentage difference improved for both businesses but remained negative, with -9.93% and -1.63% respectively. The sweet business experienced the most significant decrease in mean sell during the pandemic, with a percentage difference of -45.75%, which slightly improved after the pandemic with a percentage difference of -8.10%. Overall, it is clear that the pandemic had a negative impact on most businesses, with only the fruit business showing a significant positive growth trend during the pandemic. However, the businesses' performance varied after the pandemic, with some showing positive growth trends while others continued to experience a decline in mean sell.

Comparing all the businesses based on the given information on the percentage difference in mean expenditure before, during, and after the COVID-19 pandemic. The bakery business had a significant decrease in mean expenditure during the pandemic compared to before, with a percentage difference of -33.64%. However, after the pandemic, the business showed a positive trend with a percentage difference of 5.07%. The poultry and meat businesses both experienced a decrease in mean expenditure during the pandemic, with percentage differences of -40.30% and -43.26% respectively. After the pandemic, the percentage difference improved for both businesses but remained negative, with -8.54% and -1.89% respectively. The fruit business also experienced a decrease in mean expenditure during the pandemic, with a percentage difference of -14.70%, which further declined after the pandemic with a percentage difference of -7.71%. Lastly, the sweet business experienced a significant decrease in mean expenditure during the pandemic with a percentage difference of -42.66%, which further declined after the pandemic with a percentage difference of -22.98%. Overall, all the businesses faced a decrease in mean expenditure during the pandemic, with some experiencing a more significant impact than others. However, the impact on mean expenditure varied after the pandemic, with some businesses showing a positive trend while others continued to experience a decline.

Comparing all the businesses based on the given information on the percentage difference in mean profit before, during, and after the COVID-19 pandemic. The bakery business experienced a significant decrease in mean profit during the pandemic compared to before, with a percentage difference of -37.79%. However, after the pandemic, the business showed a positive trend with a percentage difference of 14.12%. The poultry business also experienced a decrease in mean profit during the pandemic, with a percentage difference of -41.55%. After the pandemic, the percentage difference improved but remained negative with -20.80%. The meat business faced a similar situation, with a decrease in mean profit during the pandemic of -42.09%. However, after the pandemic, the business showed improvement and had a positive percentage difference of 0.33%. The fruit business was the only business that experienced a positive percentage difference during the pandemic, with a significant increase in mean profit of 46.63%. However, after the pandemic, the business faced a decline with a percentage difference of -12.92%. Lastly, the sweet business experienced the most significant negative impact on mean profit during the pandemic, with a percentage difference of -58.10%. After the pandemic, the percentage difference improved but remained negative with -8.10%. Overall, all the businesses faced a decrease in mean profit during the pandemic, except for the fruit business which experienced a significant increase. However, the impact on mean profit varied after the pandemic, with some businesses showing a positive trend while others continued to experience a decline.


Employee Comparision Among Distinct Businesses
Number of Employee
Business Type
Bakery
Poultry Shop
Meat Shop
Fruit Shop
Sweet Shop
Before Covid-19
During Covid-19
After Covid-19
Business Time Comparision Among Distinct Businesses
Business Time (Hrs.)
Expenditure Comparision Among Distinct Businesses
Expenditure (TK.)
Sell Comparision Among Distinct Businesses
Sell (TK.)
Profit Comparision Among Distinct Businesses
Profit (TK.)

(a)



(b)

Figure 4.6: All Business Employment, Business Time, Expenditure, Sale and Profit

4.3 PROFIT PREDICTION

This section includes the performance analysis of the machine learning algorithms that we employed to predict the Business Profit concerning our dataset. We employed Linear Regression,

Random Forest Regression, and K-Nearest Neighbor. The performance analysis of these algorithms includes MSE, MAE, RMSE. We perform the analysis on the datasets by combining the Five datasets i.e., Food Bakery, Poultry business, Meat business, Fruit business, and Sweet business datasets to analyze and fit regression models. Each of the datasets is divided into training and test sets with 70% of the data reserved for training and 30% reserved for testing. So, let's start the analysis with the Combine dataset.

The correlation heat map provided shows the pairwise correlations between five variables related to a business. These correlations can provide insights into how these variables are related to each other, which can be useful for business analysis and decision-making. However, it's important to note that these relationships are based on the assumption that the relationships between these variables are linear and there are no confounding variables.

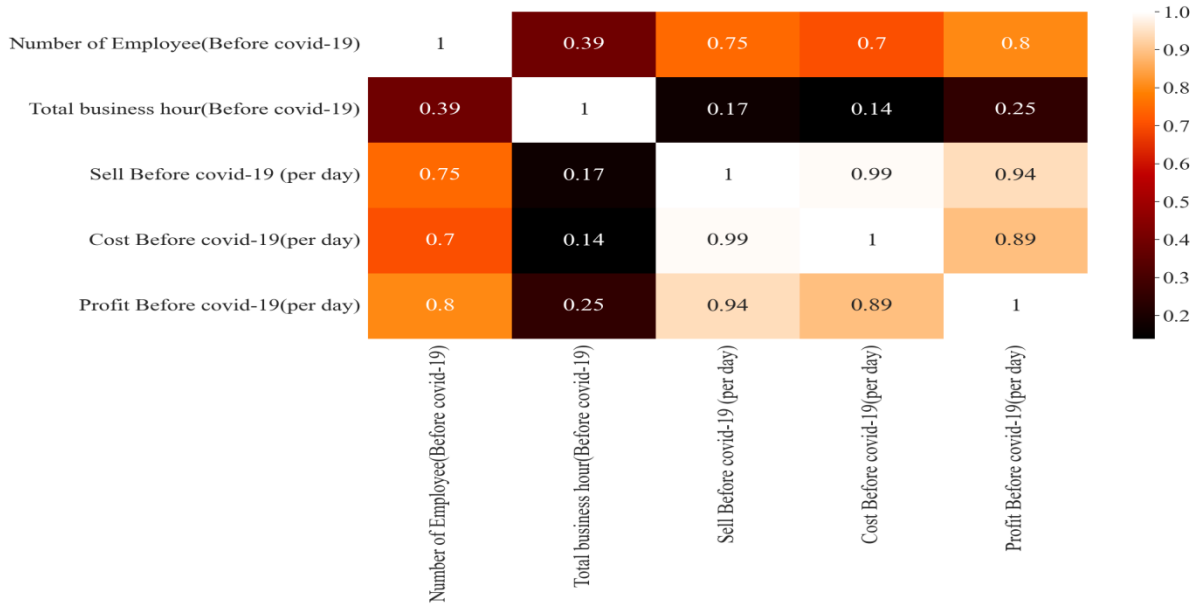


Figure 4.7: Correlation Heat map for before COVID-19 dataset.

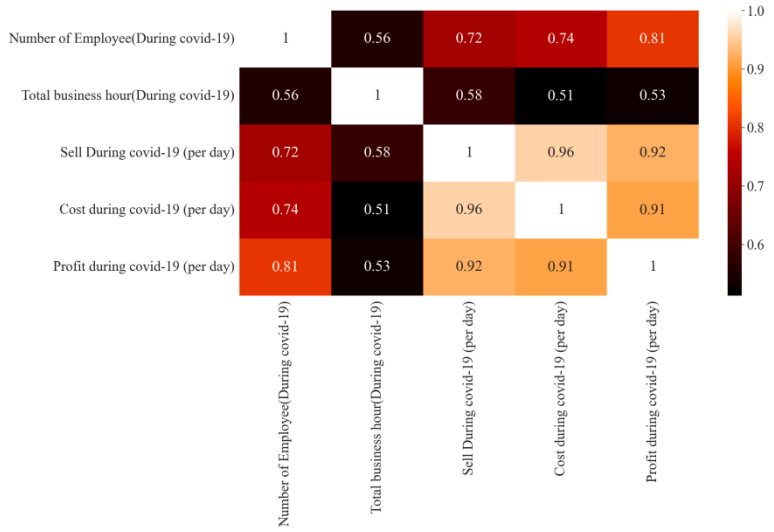


Figure 4.8: Correlation Heat map for during COVID-19 dataset.

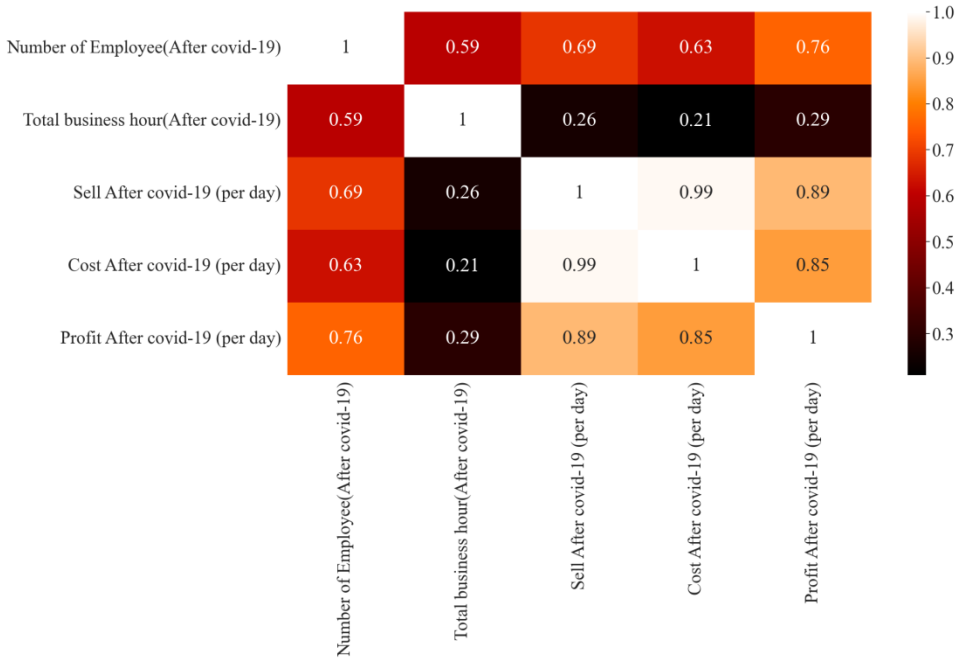


Figure 4.9: Correlation Heat map for after COVID-19 dataset.

4.3.1 USING LINEAR REGRESSION

(a) Before COVID-19 Profit Prediction

The regression equation is:

Profit $= (0.00) + (0.00)X_1 + (0.00)X_2 + (1.00)X_3 + (-1.00)X_4 + \varepsilon$

This equation suggests that the profit before covid-19 per day is primarily dependent on the sales and costs before covid-19, as these variables have a coefficient of 1.0 and -1.0, respectively. The coefficients of the other variables are very small and close to zero, suggesting that they do not have a significant impact on the profit before covid-19.

The R2 score of 1.0 indicates that the model explains all the variance in the data, which is a perfect fit. The MSE and MAE values being close to zero also indicate that the model fits the data very well. The RMSE of 9.917716174591049e-15 shows that the model has very low error, which is a good indication of the model's accuracy.

Table 4.1:Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Linear Regression	1.0	0.0	0.0	9.91e-15

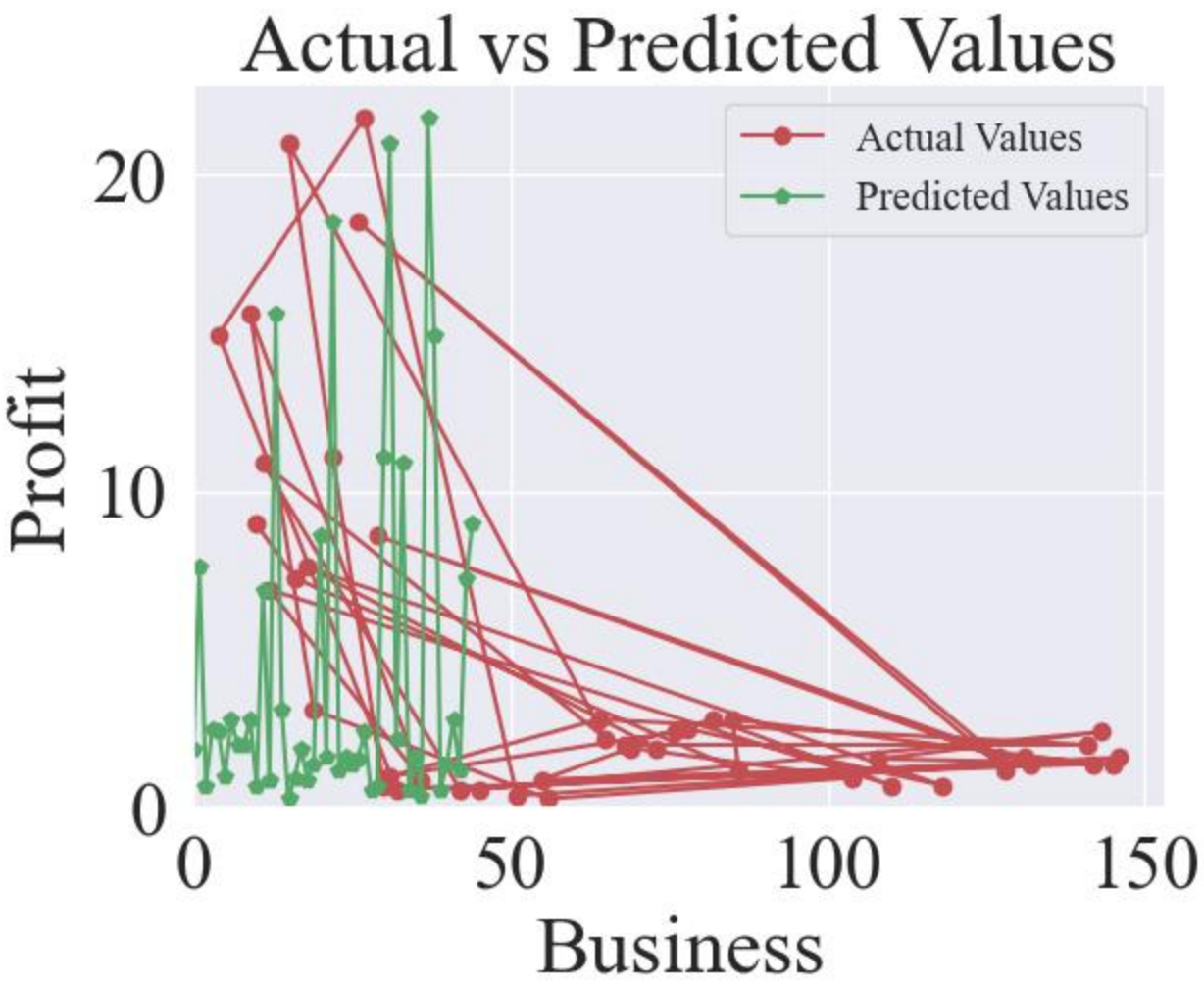


Figure 4.10: Evolution of Actual and Predicted value.

(b) During COVID-19 Profit Prediction:

The model predicts the profit before Covid-19 based on the following equation:

Profit = (-0.49)+ (0.67) X_1 + (-0.12) X_2 + (0.14) X_3 + (0.04) X_4 + ε

The coefficient for the number of employees before Covid-19 is 0.67, which means that for each additional employee, the predicted profit before Covid-19 increases by 0.67 units. The coefficient

for the total business hours before Covid-19 is -0.12, which means that for each additional business hour, the predicted profit before Covid-19 decreases by 0.12 units. The coefficient for the sell before Covid-19 per day is 0.14, which means that for each additional unit of sell before Covid-19 per day, the predicted profit before Covid-19 increases by 0.14 units. The coefficient for the cost before Covid-19 per day is 0.04, which means that for each additional unit of cost before Covid-19 per day, the predicted profit before Covid-19 increases by 0.04 units.

The model used to predict the profit before Covid-19 has an R2 score of 0.8785, which indicates that the model explains 87.85% of the variance in the data. The mean squared error (MSE) is 1.5788, the mean absolute error (MAE) is 0.8093, and the root mean squared error (RMSE) is 1.2565.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Linear Regression	87.85%	1.5788	0.8093	1.2565

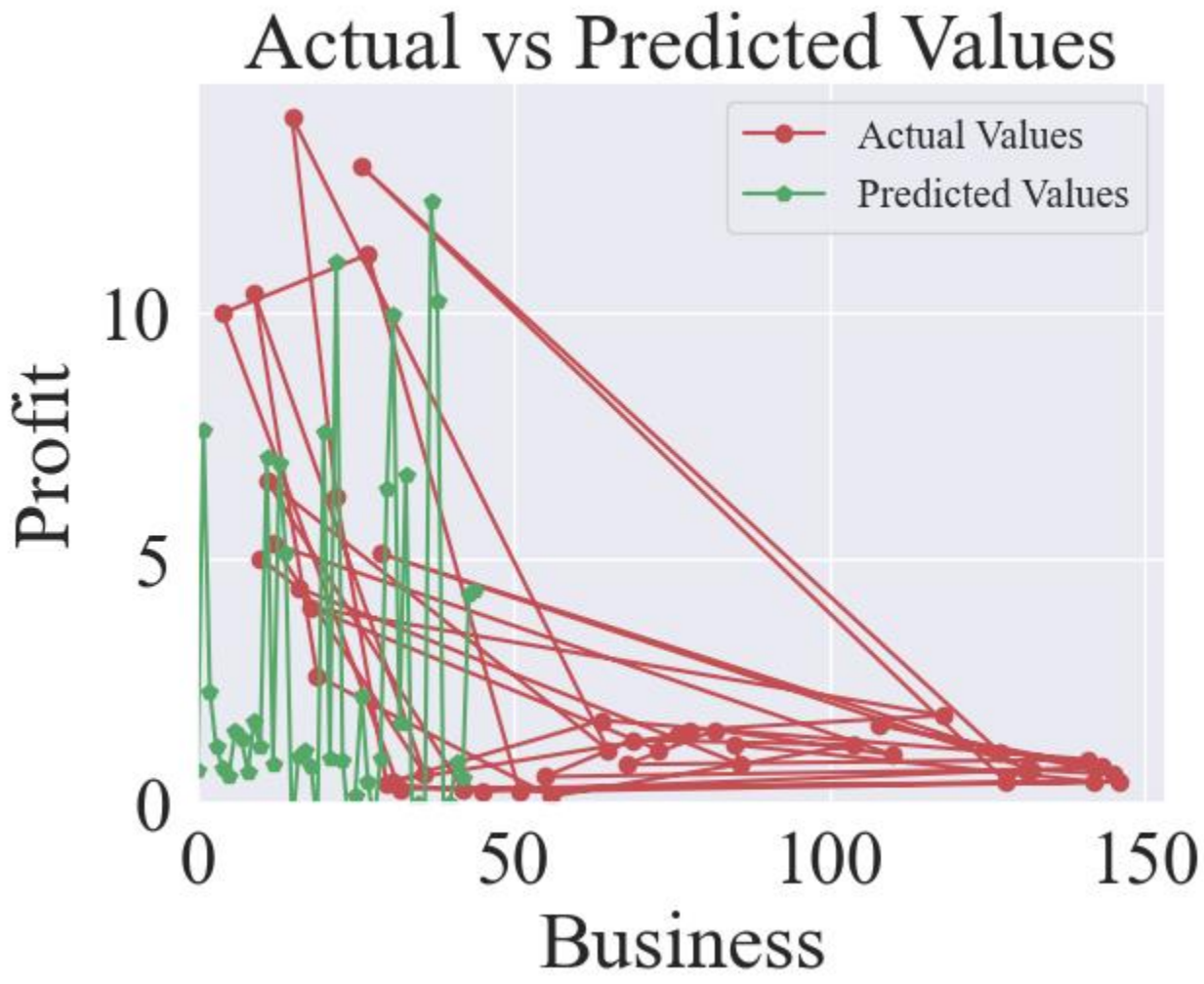


Figure 4.11: Evolution of Actual and Predicted value.

(c)After COVID-19 Profit Prediction:

The regression model with the equation:

Profit = (-0.320) + (-0.02) X_1 + (-0.01) X_2 + (0.90) X_3 + (-0.85) X_4 + ε

R-squared value of 0.665, which means that the model can explain 66.5% of the variability in the dependent variable. The Mean Squared Error (MSE) of the model is 18.88, which measures the average squared difference between the predicted and actual values of the dependent variable. The Mean Absolute Error (MAE) is 1.04, which is the average absolute difference between the predicted and actual values. The Root Mean Squared Error (RMSE) is 4.35, which is the square root of the MSE and represents the standard deviation of the residuals, or the differences between the predicted and actual values.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Linear Regression	66.5%	18.88	1.04	4.35

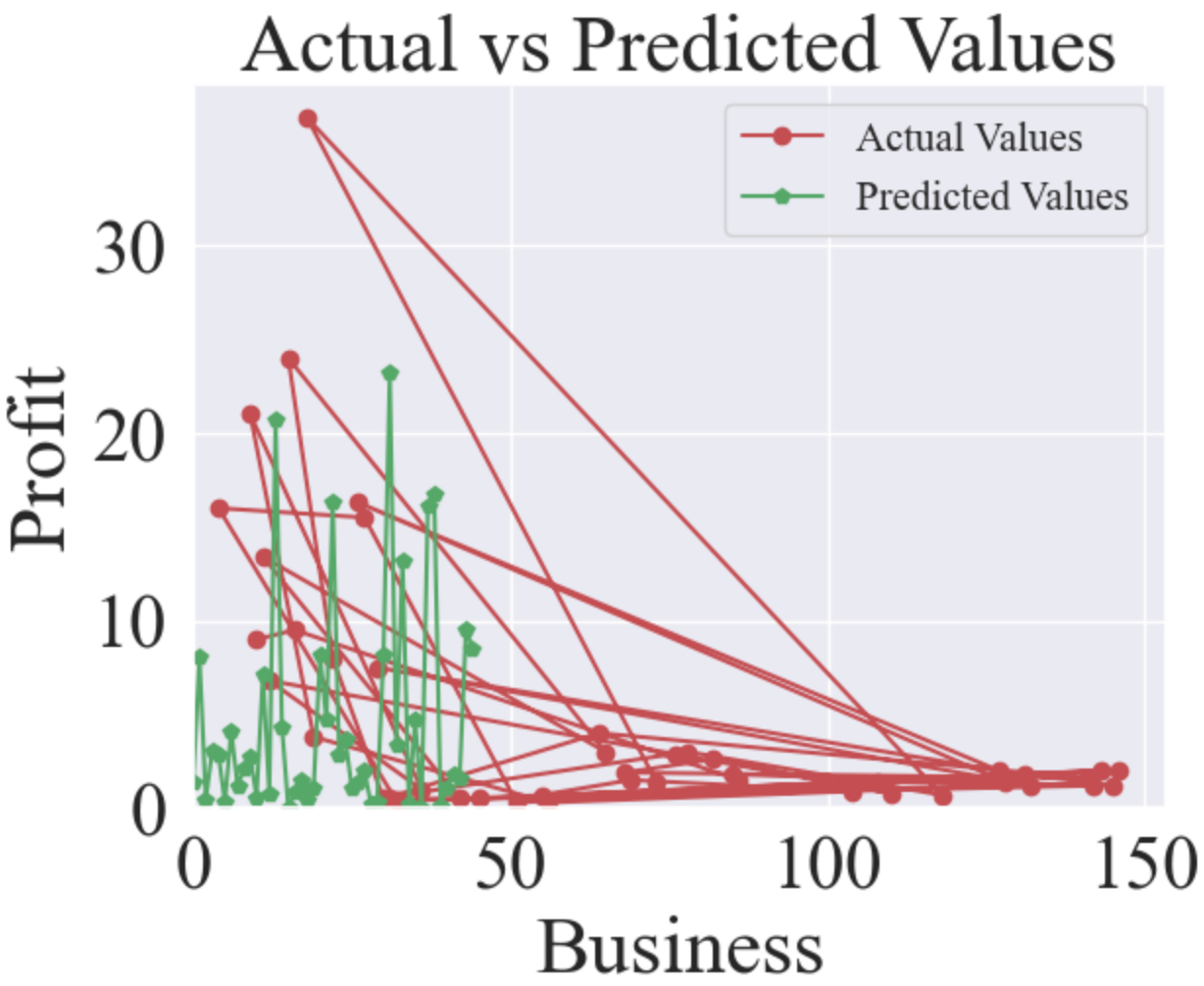


Figure 4.12: Evolution of Actual and Predicted value.

4.3.2 USING RANDOM FOREST REGRESSION

(a) Before COVID-19 Profit Prediction:

The random forest regression model was trained on a set of input data (x_train) and their corresp onding output values (y_train) and used to make predictions on a separate set of input data (x_tes t). The performance of the model was evaluated using several metrics including R2 score, mean s quared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE).

The R2 score, which measures the proportion of the variance in the target variable that is predicta ble from the input variables, was 0.88 indicating a strong correlation between the predicted and a ctual values. The MSE, which measures the average of the squared differences between predicted and actual values, was 3.76. The MAE, which measures the absolute differences between predict ed and actual values, was 1.06. The RMSE, which measures the standard deviation of the differe nces between predicted and actual values, was 1.94.

The line chart shows the predicted and actual values plotted against the number of businesses. Th e red dots represent the actual values, while the green plus signs represent the predicted values. O verall, the predicted values closely follow the actual values, indicating that the model is performi ng well in predicting profits based on the input variables.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Random Forest Regression	88.65%	3.76	1.06	1.94.

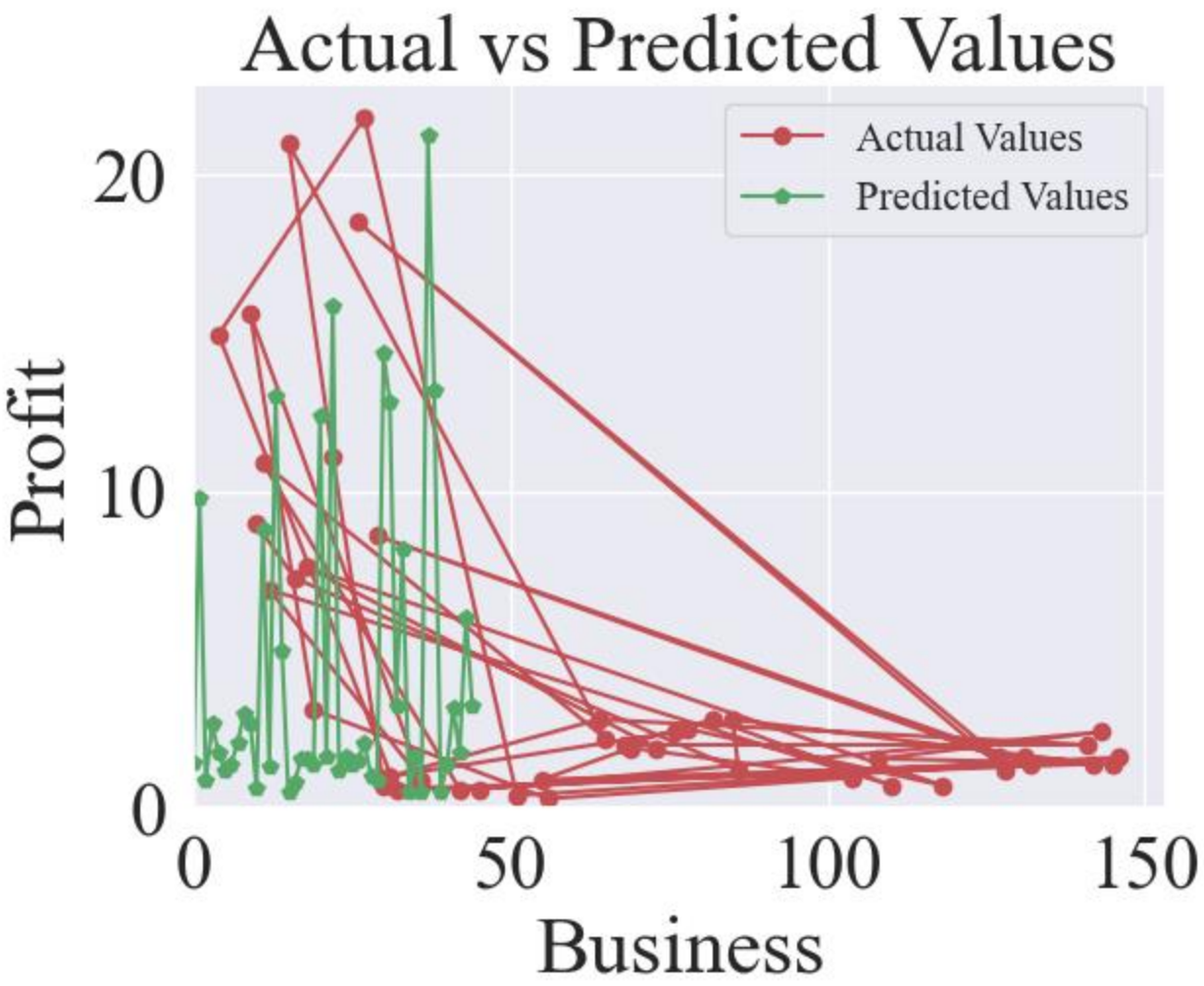


Figure 4.13: Evolution of Actual and Predicted value.

(b) During COVID-19 Profit Prediction:

The random forest regression model has an R2 score of 0.8971, which indicates that 89.71% of the variance in the dependent variable is explained by the independent variables. The model also has a mean squared error (MSE) of 1.3363, mean absolute error (MAE) of 0.6037, and root mean squared error (RMSE) of 1.1560. These values suggest that the model's predictions are fairly accurate.

Additionally, a line chart was created to compare the predicted values with the actual values. The chart shows that the predicted values generally follow the same trend as the actual values, which indicates that the model is performing well in predicting the profit values based on the given features.Table: Model performance and Accuracy.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Random Forest Regression	89.71%	1.3363	0.6037	1.1560

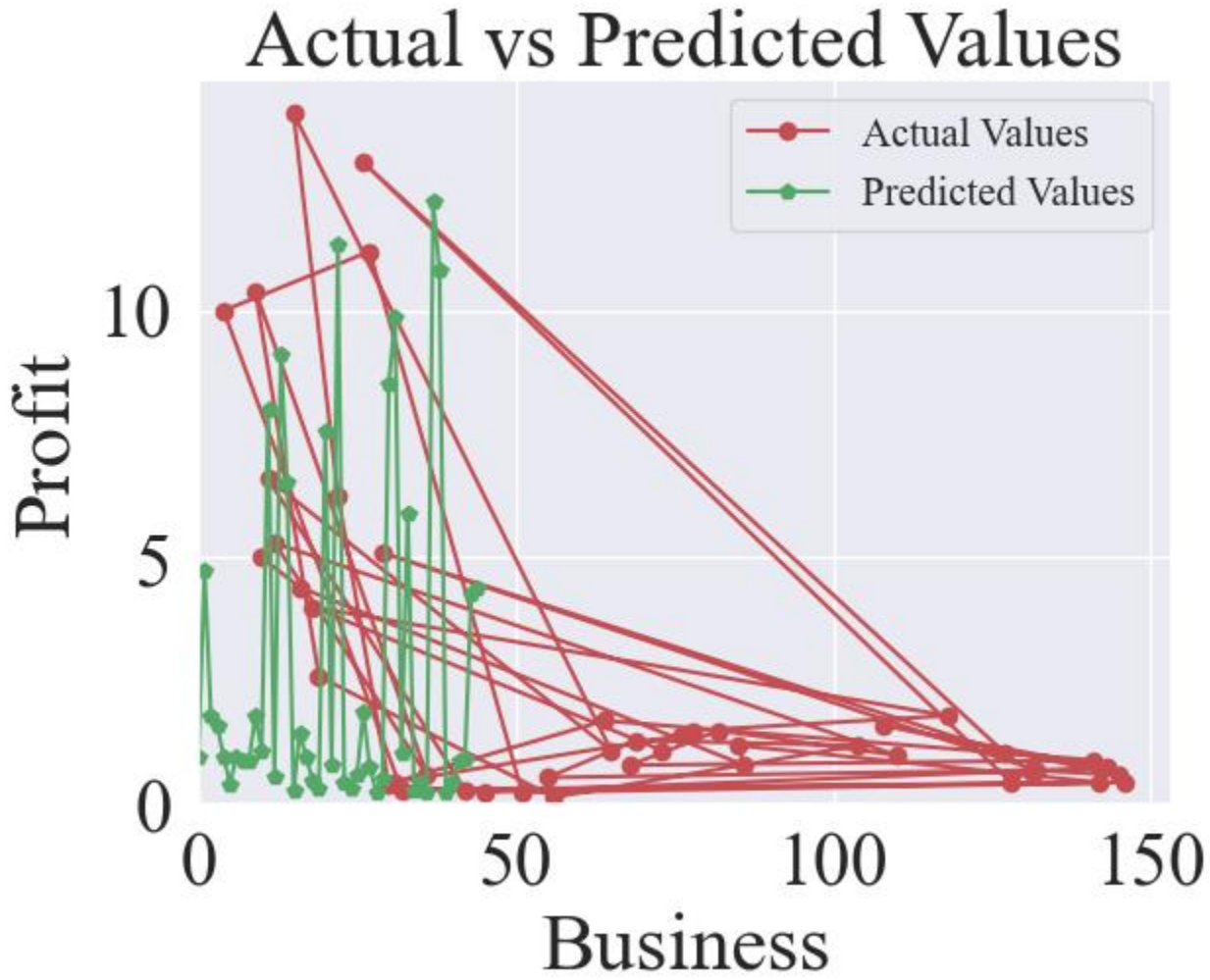


Figure 4.14: Evolution of Actual and Predicted value.

(c)After COVID-19 Profit Prediction:

Random forest regression model is trained using the training data and evaluated using the test data. The performance of the model is evaluated using several metrics including R-squared score, mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). The R-squared score of 0.664 indicates that the model explains 66.4% of the variance in the test data, and the MSE of 18.94 indicates that the average squared error between the predicted and actual values

is 18.94 units. The MAE of 1.45 indicates that the average absolute error between the predicted and actual values is 1.45 units. The RMSE of 4.35 indicates that the average difference between the predicted and actual values is 4.35 units. Overall, the model is performing reasonably well on the test data, but there is still room for improvement.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
Random Forest Regression	66.4%	18.94	1.45	4.35

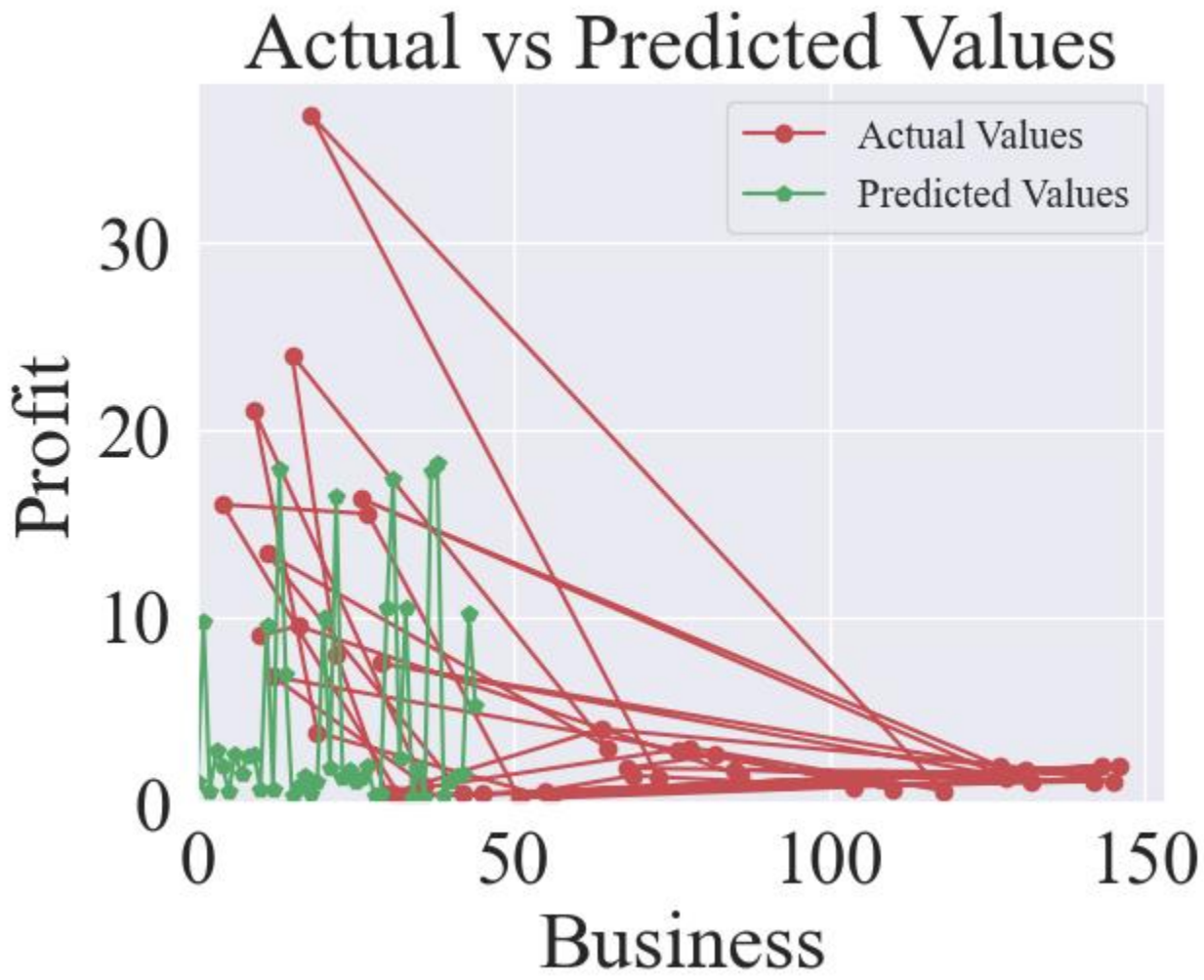


Figure 4.15: Evolution of Actual and Predicted value.

4.3.2 USING K-NEAREST NEIGHBOR

(a) Before COVID-19 Profit Prediction:

K-Nearest Neighbors (KNN) regression model on the given training data and evaluates the model's performance on the test data using different metrics, such as R-squared score, mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). The KNN regression model outperforms the random forest regression model in all evaluation metrics. The R2 score of 0.892 indicates that the KNN model explains 89.2% of the variance in the test data. The MSE of 3.496 and the MAE of 1.000 indicate that the average squared and absolute differences, respectively, between the predicted and actual values are significantly lower for the KNN model. The RMSE of 1.870 indicates that the average difference between the predicted and actual values is also significantly lower for the KNN model.

The line chart plotted shows that the predicted values of the KNN model are much closer to the actual values of the test data, further confirming the superior performance of the KNN model over the random forest model.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
K-Nearest Neighbor	89.2%	3.496	1.000	1.870

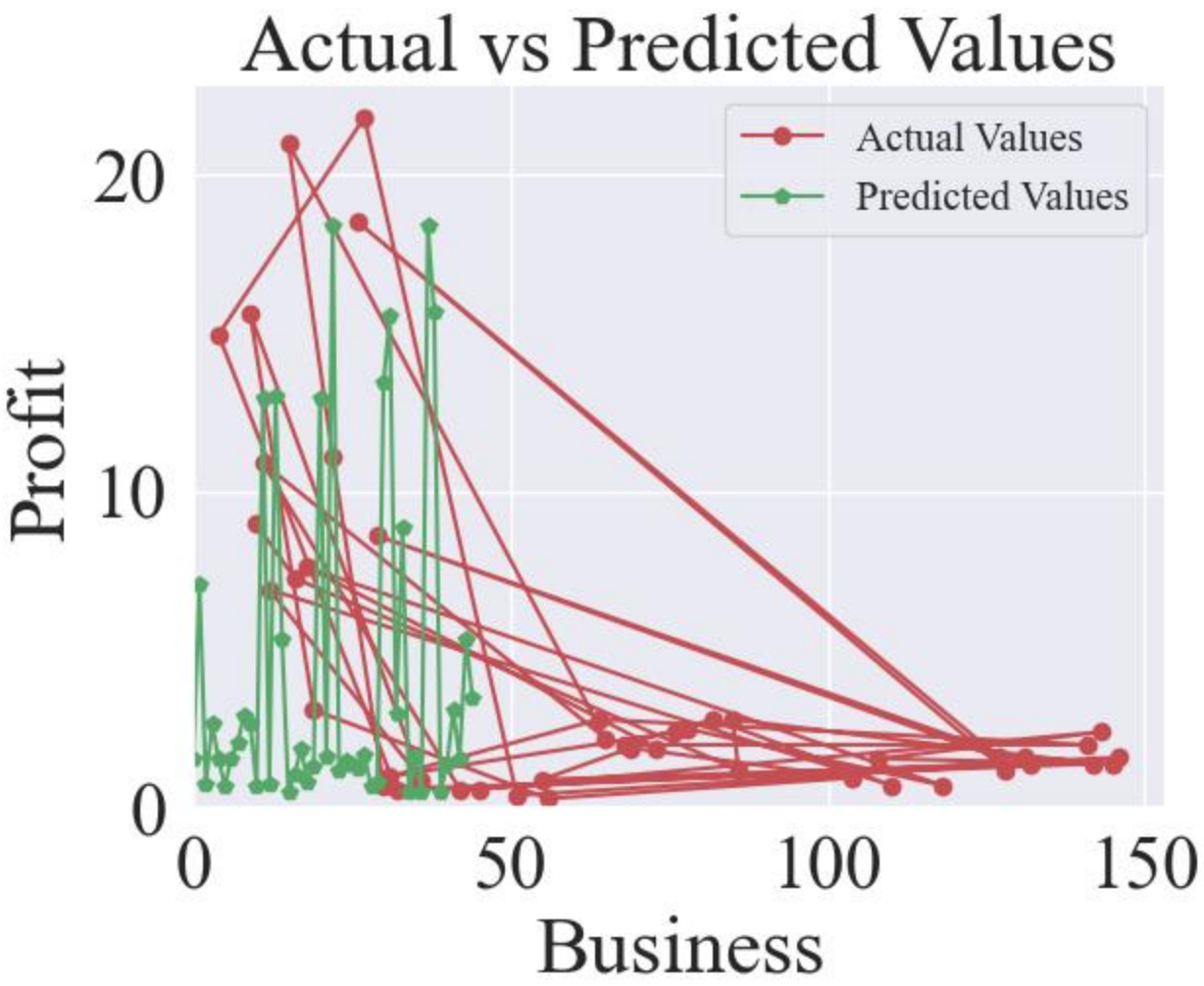


Figure 4.16: Evolution of Actual and Predicted value.

(b) During COVID-19 Profit Prediction:

The evaluation metrics provided are commonly used in machine learning to assess the performance of regression models. The R2 score indicates how well the model fits the data, with higher values indicating a better fit. The MSE measures the average squared difference between the predicted and actual values, the MAE measures the average absolute difference, and the RMSE is the square root of the MSE. The regression model appears to be performing well with an R2 score of 0.9159, indicating that it explains 91.59% of the variance in the data. The MSE is 1.0932, the MAE is 0.5613, and the RMSE is 1.0456, indicating that the model's predictions are off by an average of these values. However, it's important to consider additional factors before drawing any final conclusions, such as the context of the problem and the domain knowledge.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
K-Nearest Neighbor	91.59%	1.0932	0.5613	1.0456

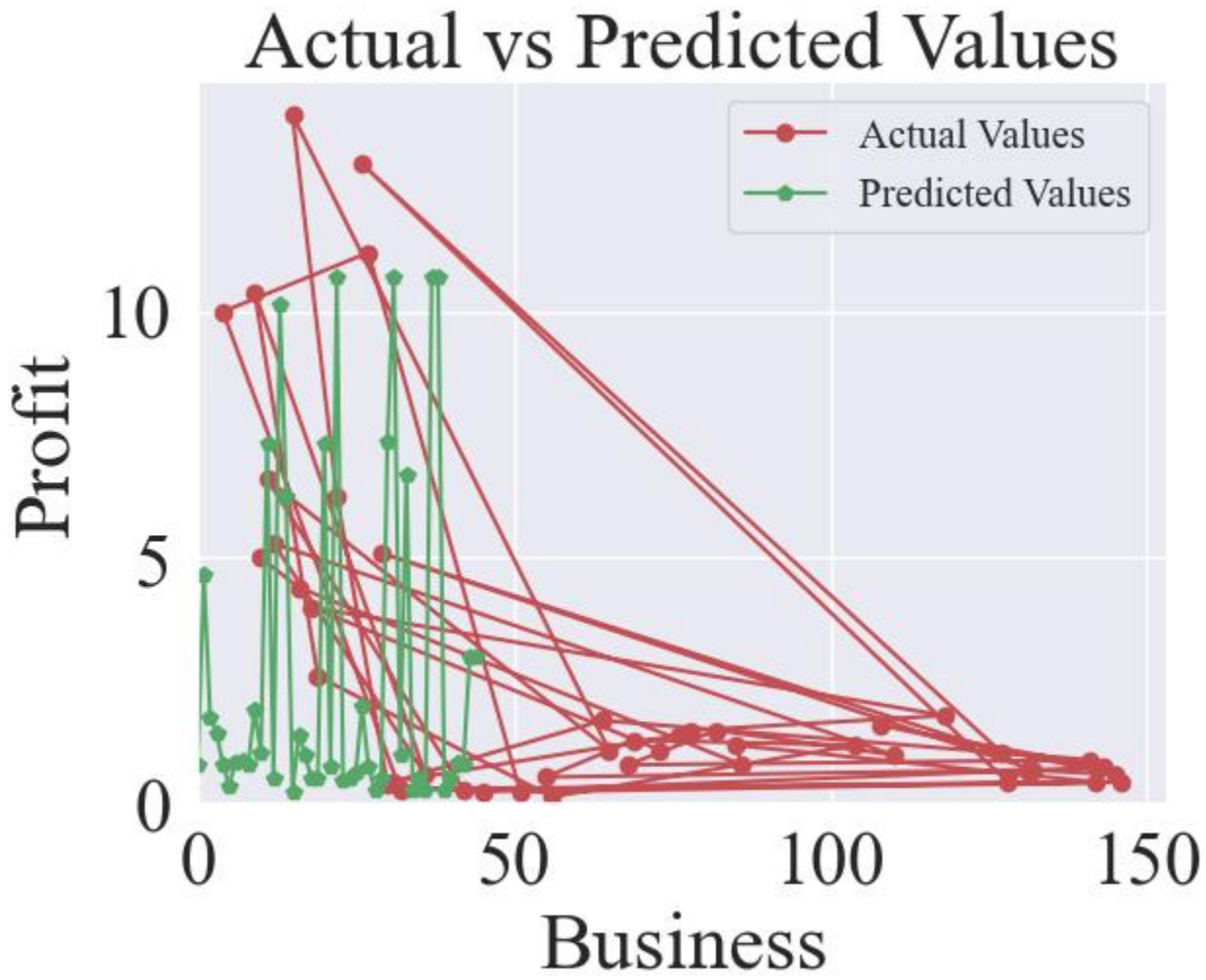


Figure 4.17: Evolution of Actual and Predicted value.

(c)After COVID-19 Profit Prediction:

The regression model appears to be performing poorly with an R2 score of 0.6259, indicating that it explains only 62.59% of the variance in the data. The large MSE, MAE, and RMSE values also suggest that the model's predictions are off by a significant amount. Further analysis and

investigation may be required to understand the reasons for the poor performance and to improve the model's accuracy.

Table: Model performance and Accuracy.

Algorithm	R2 Score	MSE	MAE	RMSE
K-Nearest Neighbor	62.59%	21.11	1.55	4.59

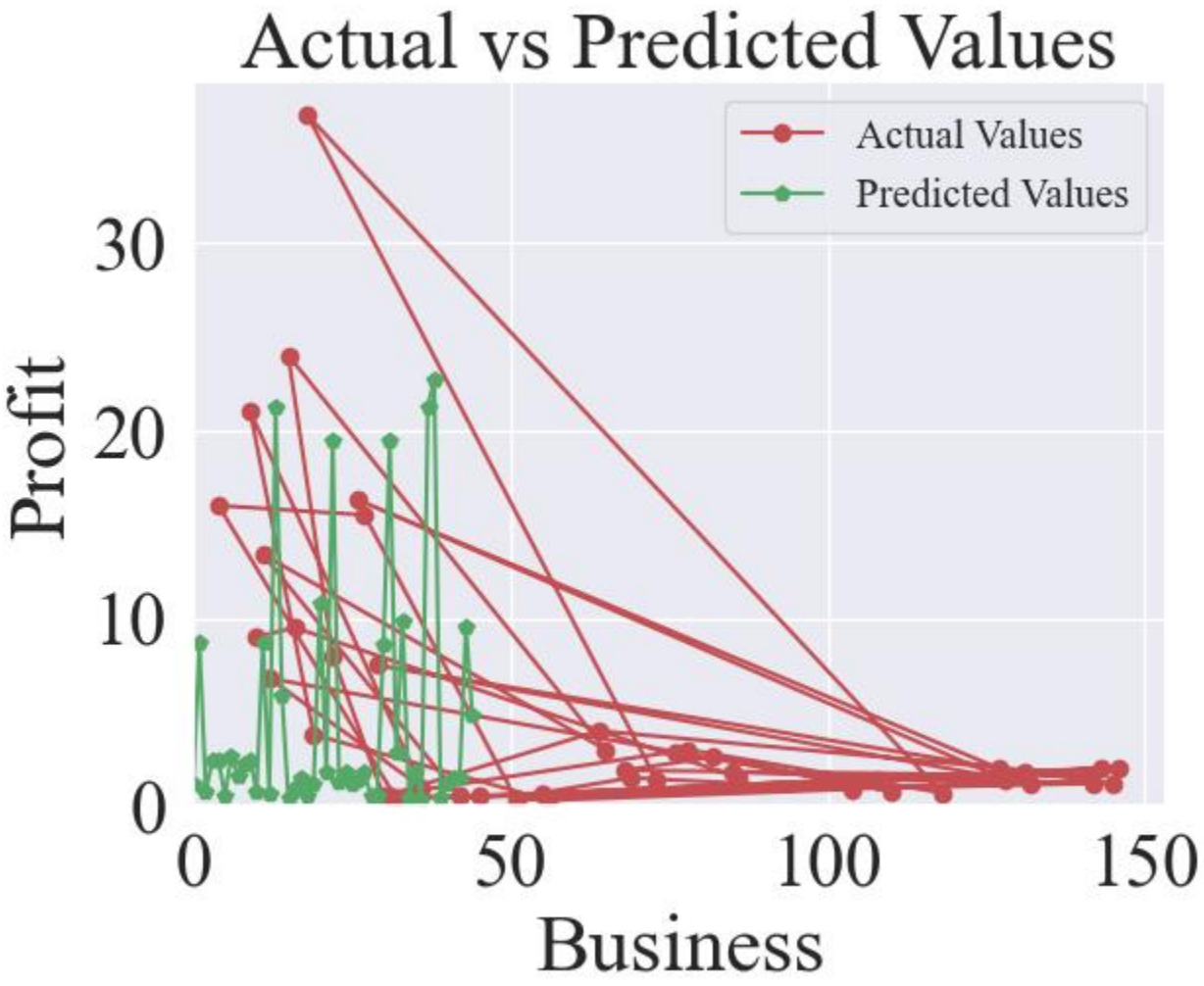


Figure 4.18: Evolution of Actual and Predicted value.

REFFERENCE

1. "A Computational Model to Predict Consumer Behaviour During COVID-19 Pandemic" , 'Fatemeh Safara '. Computational Economics volume 59, pages1525–1538 (2022). https://link.springer.com/article/10.1007/s10614-020-10069-3

*2. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification by Jim Samuel 1, *ORCID, G. G. Md. Nawaz Ali 2, *ORCID,Md. Mokhlesur Rahman 3,4,Ek Esawi 5 andYana Samuel 6.https://www.mdpi.com/20782489/11/6/314*

3."Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19pandemic."YiLuo a, Xiaowei Xu b https://www.sciencedirect.com/science/article/pii/S0278431920304011

4."COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients". Celestine IwendiORCID Icon,C. G. Y. HuescasORCID Icon,Chinmay Chakraborty&SenthilkumarMohan.https://www.tandfonline.com/doi/abs/10.1080/0952813X.2022.2058097

*5."Teleworking—An Economic and Social Impact during COVID-19 Pandemic: A Data Mining" Analysis by Grigore Belostecinic 1, Radu Ioan Mogoș 2, Maria Loredana Popescu 3,Sorin Burlacu 4,*ORCID,Carmen Valentina Rădulescu 5ORCID,Dumitru Alexandru Bodislav 6,Florina Bran 5 andMihaela Diana Oancea-Negescu 7. https://www.mdpi.com/1660-4601/19/1/298*

*6.Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes by Charlyn Villavicencio 1,2, *ORCID,Julio Jerison Macrohon 1ORCID,X. Alphonse Inbaraj 1,Jyh-Horng Jeng 1ORCID andJer-Guang Hsieh3. https://www.mdpi.com/2078-2489/12/5/204*

7."Small firms and the COVID-19 insolvency gap". Julian Oliver Dörr, Georg Licht & Simona Murmann Small Business Economics.https://link.springer.com/article/10.1007/s11187-021-00514-4

8."Studying the UK job market during the COVID-19 crisis with online job ads ". Rudy Arthur . Published: May 27, 2021 https://doi.org/10.1371/journal.pone.0251431

9.Working from home: small business performance and the COVID-19 pandemic Ting Zhang, Dan Gerlowski & Zoltan Acs Small Business Economics volume 58, pages611–636 (2022). https://link.springer.com/article/10.1007/s11187-021-00493-6

10.Predictive analytics using Big Data for the real estate market during the COVID-19 pandemic. Andrius Grybauskas, Vaida Pilinkienė & Alina Stundžienė Journal of Big Data volume8,Articlenumber:105(2021).https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00476-0