

Supershop management Models: An optimised way to
Manage the supershop Using Hyperautomation and Machine
Learning

by

Shuvro Ahmed

18201119

Rajesh Mojumder

18201170

MD.Mahmudur Rahman

19101119

Joy Karmoker

19301271

Shadman Fatin

22241128

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

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
23 January 2023

© 2023. Brac University
All rights reserved.

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Shuvro Ahmed
18201119

Rajesh Mojunder
18201170

MD.Mahmudur Rahman
19101119

Joy Karmoker
19301271

Shadman Fatin
22241128

Approval

The thesis/project titled “Supershop management Models: An optimised way to Manage the supershop Using Hyperautomation and Machine Learning” submitted by

1. Shuvro Ahmed (18201119)
2. Rajesh Mojumder (18201170)
3. MD.Mahmudur Rahman (19101119)
4. Joy Karmoker (19301271)
5. Shadman Fatin (22241128)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on 23 January 2023.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Md Tanzim Reza
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

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

Abstract

Customers are the heart and soul of supermarkets, and these stores often suffer losses due to mishandling of customer service. This study aims to examine how the satisfaction of customers can be maximized with the help of hyper-automation technologies in order to operate the stores successfully. Here, an intelligent voice bot is used to reduce response time for basic customer queries by providing real-time replies using NLP and for further complex queries, customers will be provided with contact info or forwarded to relevant authorities. In addition to that we predicted customer demands for future products by using multiple machine learning libraries like XGBoost, Linear Regression and Random forest with the help of daily sales data. This helps supermarkets to stock a perfect amount of products in their inventory which will help them to avoid any kind of product shortage in any season and will help them to achieve customer satisfaction. Furthermore, optimized product placement will be ensured with the use of data mining techniques like Apriori algorithm, FP Growth algorithm and GSP algorithm. By doing this we are making it easy for the customers to find out their preferred product together in a single shelf and reducing customer hassle of iterating through the whole shop to find the products from their shopping list. As well as, to ensure a hassle-free transaction between consumer and seller, a system is proposed using Smart Contract System via Block chain which can make the process faster, and ensure transaction safety at the same time. For results, With an R-Squared score of 0.963, we discovered that the hybridization of linear-boost regression was the most suited for forecasting. The best outcomes for product placement were provided by FP Growth. For the purposes of the chatbot, let's say that for the two strings "rfl nipple 3-6 month" and "rfl nipple 3 to 6 month," the spaCy library and nltk's bleu function both yield 90.8 and 66.21 percent similarity, respectively. Now, based on the %, you could assume that the spaCy library is operating more effectively, but this is untrue. SpaCy library functions. better in a big model where pre-trained word vectors are present, but not in a small model. However, the smart contract system successfully carried out all of the system's algorithms and guaranteed transaction security as well as product safety.

By using the Hyper Automation technology Super stores can ensure better customer service. The budget and implementation of such technologies combined into a system turned out to be both high and complicated. However, as time passes, the cost of these technologies will decrease rapidly and their usage will be further simplified.

Keywords: Hyperautomation, Data Mining, Machine Learning, NLP, Smart Contact system, Voice Bot, spaCy, Linear Regression Analysis, Time Series Analysis

Acknowledgement

Firstly, all praise to the almighty God for whom our thesis have been completed without any major interruption.

Secondly, We are grateful to our Supervisor Md. Golam Rabiul Alam, PhD and Co-supervisor Md Tanzim Reza sir for his kind support and advice in our work. They helped us whenever we needed help.

Moreover, special Thanks to Md.Tawhid Anwar Sir with all other seniors, batch-mates, friends and juniors who helped in different aspects in our thesis work.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration	i
Approval	ii
Abstract	iv
Acknowledgment	v
Table of Contents	vi
List of Figures	viii
List of Tables	1
1 Introduction	2
2 Literature Review	5
2.1 Background Study	5
2.2 Related Works	10
2.2.1 A literature review of chatbot and voice recognition systems .	10
2.2.2 Literature review on product placement	11
2.2.3 Literature review of product sales forecasting	14
2.2.4 Literature review on Smart Contract System	15
3 Methodology	17
3.1 System Overview	17
3.1.1 System Diagram	17
3.1.2 System Description	18
3.2 Dataset Overview	19
3.2.1 Data Description	19
3.2.2 Data Preprocessing	19
3.3 Model	21
3.3.1 Voice Chat Bot Methodology and Implementation	21
3.3.2 Product Placement Methodology and Implementation	23
3.3.3 Product Sales Forecasting and Implementation	25
3.3.4 BLOCKCHAIN: Smart Contract System process and Imple- mentation	28
4 Result Analysis	32

5 Conclusion	40
5.0.1 Future Work	40
Bibliography	43

List of Figures

3.1	Supershop Management System Diagram	17
3.2	Product Sales Time Plotting	20
3.3	Moving Average Graph of the Entire Dataset	20
3.4	Linear Trend Generation	21
3.5	The concepts of Voice Chat Bot	21
3.6	Flowchart Diagram of a general Voice Transformation	22
3.7	Workflow Diagram of Product placement Methodology	25
3.8	Workflow of Forecasting	26
3.9	Linear Trend forecast	27
3.10	Workflow Diagram of Smart Contract System Transactions Method	29
4.1	product category performance	34
4.2	product name performance	34
4.3	Forecasting using linear regression	36
4.4	Forecasting using random forest	37
4.5	Forecasting using xgboost	37
4.6	Forecasting using linear regression-random forest	38
4.7	Forecasting using linear regression-xgboost	38
4.8	Successful Transaction	39

List of Tables

4.1	Product Category	33
4.2	Product Name	33
4.3	Performance of Different Machine learning Models	36

Chapter 1

Introduction

Introduction

Hyper Automation is a system that works with different sectors of artificial intelligence, such as robotic process automation, machine learning, etc. to make our lives easier. As it replaces humans with artificial intelligence, it has the ability to give a boost to different business sectors. The human body needs rest after working for a certain amount of time. But machine-based systems can work 24/7 without any rest, which can double the production rate. In hyper-automation, the system first recognizes which parts of that particular sector can be automated. After that, it works on finding suitable AI or machine learning techniques that can be used to increase the capability of working. This hyper-automation allows enterprises to expand. It is a noticeable fact that after the pandemic people are getting more reliable in hyper-automation. The main reason for this reliability is that it helps organizations to execute tasks at a faster speed with much more accuracy at a lower price. In this paper, we are mainly focusing on 5 parts or components, which will form the basis of our framework. These parts are: chat and voice bot for providing faster responses to queries, generating a future prediction of the number of products sold in the next month or next day comparing all the monthly data, crowd management, product placement and. These features or parts will mainly use machine learning models to generate predictions or NLP to process text. They will also use data mining to predict product placement. A job queue will also be used for crowd management. A brief summary of these terms is given below: As a sub-field of artificial intelligence, machine learning defines the process by which computers learn to recognize patterns in data and make predictions based on that given data by analyzing the previous sales record it is possible to make an estimation on how the seller can go in future days. Besides, Natural Language Processing (NLP) is the automated manipulation of natural language by software, such as speech and text. A company's data mining process transforms unstructured data into useful information. Using software to look for patterns in large amounts of data, companies may gain a better understanding of their consumers. So they can create better marketing strategies, increase sales, and lower expenses. In order to data mining to be successful, it is necessary to have well-functioning data gathered and processed. For the payment system is also updated with the touch of the Blockchain model. We have sued a smart contract system through which customers can process their payment faster with less risk of any security issues while processing the payment while the super

shop owner can get the transaction money in time.

In order to provide excellent customer service, business owners must build trusting connections with their customers. By making them feel valued as a customer, you'll guarantee that they leave your shop with a great image of your business. Customer loyalty and favorable social media evaluations will follow from this strategy.

A. Motivation

In our research, our main motivation was to enrich the customer experience and help the supershop owners to grow more profit in the business. In Bangladesh, we still do not have automated management system which results in hassle for customers and sellers alike. Especially during rush hour and in crisis, this is a severe problem. We have seen during the covid pandemic, super shops had to close down or limit service due to being managed fully by human employees. But having an automated system with a faster response time can solve these problems easily. Moreover, another purpose of the research is to avoid the fraud of transactions and ensure the Security of the product. Mainly the system used in this research will solve the general super shop problems in a developing country.

B. Research Problem

In modern days, in general, super shops face several types of problems, such as the attraction of new customers, the faith of new customers and retaining them, and trying to figure out the customer's expectations, market competition, and a variety of payment methods. The application of hyper automation can solve all these problems in a super shop and can easily make a good amount of profit for the super shop owner. Smart superstores employ IoT sensors to capture physical data, transmit this data via digital platforms to the grocery store workers, and assist them in making better decisions. Jean-Michel Finally, a digital strategy consultant at IBM, says that "IoT in supermarkets is essentially about linking sensor technology with goods that are regularly utilized in the shop." The pricing point has shifted recently, though, as devices have become cheaper and more cost-effective, so the juice is now worth the squeeze, according to him. There are several features of a smart superstore, such as contact less technologies, smart shelving, and AI-based chat bots instead of human staff. Businesses will benefit from a more efficient superstore. When compared to the current superstore model, this will be a more cost-effective solution. As a result, not only will customers have a better experience, but a large number of clients will be easier to handle. Smart systems in supermarkets are now more accessible than ever before, thanks to the falling cost of technological gadgets.

C. Research objectives

Products, prices, and store environments are all factors that contribute to a positive in-store experience. However, it is the level of customer service that will be remembered. And in the age of viruses that force social distancing, lack of time, and abundance of technology, every customer wants the optimized version while shopping. While natural & professional human interactions can always lead to a friendly customer, it is only through quick response & hassle-free shopping experience that will turn the customer into a loyal one. Through the usage of hyper-automation technologies, we can easily build a healthy customer-seller relationship while making sure both parties get the optimized ROI. This will also allow people to save precious time. And most importantly, while our proposed framework may be primarily used for supermarket optimization, it can be easily modified to be used for any individual seller or startup looking to sell or pitch their products in an effective way. It will also mean fewer human errors, leading to a better experience for everyone involved.

Chapter 2

Literature Review

2.1 Background Study

The main objective of this research is to find and establish an optimized system which can ensure customer satisfaction and in automated way. So, the first step for that is creating a voice bot. Computer programs known as voicebots and chatbots are made to mimic human communication using speech or text input. They are employed for a number of purposes, including customer service, virtual assistants, and entertainment. Understanding and responding to natural language input is a major challenge in the creation of voicebots and chatbots. Natural language processing (NLP) methods like machine learning and deep learning are required for this. Making the discussion as human-like as feasible is another issue, which requires the ability to generate speech that sounds natural and to understand the context and purpose of the conversation. The purpose of research in this field is to develop NLP techniques that are more complex and to improve the ability of voicebots and chatbots to comprehend and interact with their surroundings. In addition, a number of academics are looking at ways to customize voicebots and chatbots, which will allow them to better perceive the diverse requirements of users and provide better service.

The development of voicebots is dependent on the primary technologies of machine learning and natural language processing (NLP). If you want to continue developing a voicebot, there are several steps you can take, including the following:

- 1) Become knowledgeable about machine learning and natural language processing (NLP). When developing voicebots, it is essential to have a solid understanding of both how they function and how to make use of them.
- 2) Educate yourself on the many platforms and technologies that are accessible for the production of voicebots: The creation of voicebots is feasible because to the existence of a variety of platforms and technologies, such as the Microsoft Bot Framework, Amazon Lex, and Google Dialogflow.
- 3) Collecting and annotating data: In order to train your voicebot, you will need a substantial volume of data that has been correctly labeled with the desired goal or action.
- 4) Get your voicebot ready and put it through its paces: When you've finished gathering your data, you can put it to use by teaching your voicebot and gauging how well it performs.
- 5) Implement and deploy your voicebot: Once you have tested your voicebot and

made any necessary adjustments, you are ready to launch it on a platform such as Google Assistant or Amazon Alexa.

When it comes to creating a voicebot, there are a few limitations that you could run against, such as the following:

- 1) Limited data availability: In the absence of sufficient data, teaching your voicebot to grasp and respond to input in natural language may prove to be tough.
- 2) The intricacy of natural language: Due to the fact that human language is so complicated, it might be difficult to educate a computer to grasp it and react to it in a natural way.
- 3) Voicebots may have difficulty understanding the context and meaning of a user's statements because of their limited awareness of the context and purpose of the user's remarks.

The following are some methods that might be utilized in the creation of a voicebot: Utilize pre-trained models such as BERT or GPT-3 to rapidly develop a voicebot with a high level of natural language comprehension. This may be accomplished by utilizing pre-trained models.

Transfer learning involves beginning with models that have already been pre-trained and then refining those models by utilizing one's own data in order to improve performance.

Models that are part of an ensemble can be combined in order to improve performance and reduce the number of errors that occur.

Cloud-based services: If you want to simplify the process of installing and expanding your voicebot, you may make use of cloud-based services such as Amazon Web Services or Google Cloud.

The next step is making it simpler for clients to obtain the things they want, and this is where product placement comes into play. The concept of product placement in supermarkets stems from the idea that products that are easily accessible and within the range of browsing customers are more likely to be sold. This, in turn, raises more profit for the seller and improves overall customer satisfaction. So, to achieve optimized customer satisfaction, automation of product placement is a significant part of the overall process. To achieve that, a machine learning framework is used to identify customer behavior patterns, mainly using association rule mining or ARM. Everyone in this world is having a busy life and to make our life easier we try our best to save some time from our daily work. Daily shopping is one of the most important parts of our day to day life and it is not possible to spend a huge amount of time on that. If someone wants to buy two products only and has to roam around the shop for finding the second product after buying the first product then it is consuming too much time. By the use of data mining algorithms we have tried to find out the patterns of buying products of customers and based on that we are trying to generate sequential rules. Our goal is to place the products according to the rules. Through this we are ensuring that the products that the consumer buys together usually are kept in different shelves but the shelves are kept together. And because of this the consumer doesn't have to spend too much time looking for things.

After that forecasting comes into play as this enables business to gain insights into customer's needs. Forecasting of product sales or any financial insights is required to understand the future trend of business. As, this is a thesis about optimization

of customer experience, by using the forecasting the models, one can have better understanding of customer needs and also a general idea about the profit of any products. This in turn allows a smoother customer experience especially during crisis.

Time series analysis is a statistical technique that is used to analyze, model and forecast time-dependent data. This data typically includes observations collected at regular intervals, such as hourly, daily, weekly, or monthly. The goal of time series analysis is to understand the underlying patterns, trends and seasonality in the data, and to use this information to make predictions about future observations.

There are several methods used in time series analysis, including:

Descriptive statistics: This is used to summarize the main characteristics of the data, such as the mean, median, and standard deviation.

Visualization: Time series data is often plotted as a line graph to help identify patterns, trends, and outliers.

Stationarity: A time series is said to be stationary if its statistical properties do not change over time. Stationarity is an important assumption in many time series models.

Decomposition: Time series data can be decomposed into its trend, seasonal and residual components.

Time series forecasting: There are several methods for forecasting time series data, including simple moving averages, exponential smoothing, and ARIMA (Autoregressive Integrated Moving Average) models.

Time series modeling: This is used to identify the underlying patterns and relationships in the data, and to make predictions about future observations. It includes techniques such as ARIMA, SARIMA, VAR, VARIMA, and state-space models.

Time series anomaly detection: This is used to identify unusual or unexpected observations in the data, which could indicate an error or an interesting event.

Overall, Time series analysis is an important tool in various fields like finance, economics, manufacturing, engineering, and many more. It helps to understand how the data evolves over time and make predictions about future observations.

Besides, blockchain based system enable us to have smooth and secured transactions. A computer software known as a "smart contract" is a component of blockchain technology that, when specific requirements are satisfied, causes the terms of a contract to be automatically carried out. Because they are encrypted, public, and impossible to change, these contracts are kept and copied on a distributed ledger known as a blockchain. They have a wide range of applications, including financial transactions, the management of supply chains, and the verification of digital identities, among others. Although smart contract systems are most commonly linked with the Ethereum blockchain, its implementation is not limited to that particular

platform and is possible on other blockchains as well.

Importance of using Smart contracts:

Transparency: Smart contracts are transparent, meaning that all parties involved can see the terms of the contract and the execution of it.

Flexibility: Smart contracts can be customized to meet the specific needs of a particular application, and can be used for a wide range of purposes, such as financial transactions, supply chain management, and digital identity verification.

Cost-effective: Smart contracts can reduce the need for intermediaries and legal costs, resulting in lower transaction fees.

Accessibility: Smart contract functionality can be integrated into decentralized applications, making it more accessible to users.

In a blockchain network, a smart contract is stored on the network as a digital transaction and is executed automatically when certain conditions are met. The conditions and terms of the contract are encoded in the form of code, which is stored on the blockchain and can be seen by all the participants in the network. Once the conditions of the contract are met, the smart contract automatically executes the pre-defined actions specified in the code, such as transferring funds or updating a digital asset. For example, when a user initiates a transaction to transfer funds to another user, the smart contract checks if the user has sufficient funds and if the recipient's address is valid. If both conditions are met, the smart contract automatically executes the transfer of funds to the recipient's account. Because smart contracts are stored on a blockchain, they are transparent, immutable and secure, ensuring that the terms of the contract are executed as intended, and that the record of the transaction is permanent.

The working process of smart contracts on a blockchain can be broken down into the following steps:

Creation: A developer writes the code for the smart contract, which defines the terms and conditions of the contract and the actions to be executed when those conditions are met.

Deployment: The smart contract code is deployed to the blockchain network by the developer. This step typically involves paying a fee to the network, called "gas" in the Ethereum network, to cover the computational costs of executing the contract.

Execution: Once the smart contract is deployed, it can be triggered by any user who meets the conditions specified in the contract. For example, a user may initiate a transaction to transfer funds to another user, which triggers the smart contract to check if the user has sufficient funds and if the recipient's address is valid.

Verification: The smart contract's code is executed and the conditions specified in the contract are verified by the network's nodes. If the conditions are met, the con-

tract's action are executed, for example, transfer of funds or update of a digital asset.

Recording: The outcome of the smart contract execution is recorded on the blockchain, providing an immutable and transparent record of the transaction. **Access:** The smart contract code and its execution history can be accessed by any participant on the blockchain network, providing transparency and trust in the contract's execution.

Personal BlockChain Network-

Ganache is a personal blockchain for Ethereum development. It allows developers to quickly set up a personal Ethereum blockchain for testing and experimentation purposes. It can run on a local machine or as a virtual machine and is available as a command-line tool or as a user-friendly graphical interface. Ganache provides a range of features for Ethereum developers, including the ability to: Create a virtual Ethereum blockchain that runs locally on a developer's machine Automatically generate a set of test accounts with pre-funded Ether Control the mining process to test different scenarios Inspect and analyze the state of the blockchain at any given point in time Easily import and export blockchain data Ganache is particularly useful for testing and debugging smart contracts, as it allows developers to test their contracts in a realistic environment without the need for a live network. This can help identify and fix bugs and errors before deploying contracts to the main Ethereum network.

Truffle framework-

Truffle is a development framework for Ethereum-based decentralized applications (dapps). It is an open-source tool that aims to make it easier for developers to build, test, and deploy smart contracts on the Ethereum blockchain. It's a development environment, testing framework and asset pipeline for Ethereum.

Truffle provides a suite of tools for Ethereum development, including:

- 1.A development environment for building and testing smart contracts using Solidity, Vyper or other programming languages.
- 2.A command-line interface for interacting with the Ethereum network, such as deploying and managing smart contracts.
- 3.A built-in contract compilation and deployment system, that allows you to compile, deploy and link smart contracts easily and efficiently.
- 4.A library of smart contract patterns and best practices, that can be used as a reference for developers.
- 5.A built-in testing framework for automated testing of smart contracts, that allows developers to test their contracts before deploying them to the main Ethereum network.
- 6.A built-in contract abstraction layer for easier contract interaction, that makes it easier to interact with deployed smart contracts.

7.A An integration with popular development frameworks like Ganache, which allows you to use Truffle with a local blockchain for testing and development purposes. Moreover, truffle is a widely used framework in the Ethereum ecosystem, it helps developers to streamline the development process and improve the quality of their smart contracts, making it more secure, reliable and efficient.

Solidity-

Solidity is a programming language specifically designed for writing smart contracts on the Ethereum blockchain. It is used to create decentralized applications (dapps) that run on the Ethereum Virtual Machine (EVM). Smart contracts are self-executing contracts with the terms of the agreement written directly into code. They are stored and replicated on the blockchain, making them secure, transparent, and tamper-proof. The use of Solidity in blockchain allows for the creation of decentralized apps that are able to execute complex logic and interactions, such as financial transactions, supply chain management and digital identity verification. Smart contracts in Solidity can be customized to meet the specific needs of a particular application, and can be used for a wide range of purposes. One of the main advantages of using Solidity in blockchain is that it allows for trustless interactions between parties. Smart contracts are self-executing and the terms of the contract are encoded in the code, eliminating the need for intermediaries and reducing the risk of fraud. This can also help to reduce transaction costs, as intermediaries are often a major source of fees in traditional transactions. Additionally, Solidity is a statically typed language which helps to prevent common coding errors, and makes it easier to write secure smart contracts. The use of Solidity in blockchain helps to ensure that the contracts are functioning as intended before they are deployed to the main Ethereum network, and it's also a widely used language in the Ethereum ecosystem, making it easier for developers to find resources and support.

2.2 Related Works

2.2.1 A literature review of chatbot and voice recognition systems

Language Processing (NLP) is used by the bot to connect with human users in a conversational manner. It is possible for the user to communicate with the bot in a variety of different ways. When using 'Doly,' for example, the number of affirmations it can respond to and the accuracy of each answer increase their relationship to the input affirmation [1]. Normally, users can chat with chatbots using any format, and there is no specific format that the users have to follow. In most cases, the answers given by the chatbots are correct and valid [2]. With the use of AI chatbots, the number of employees of an organization can be reduced by a huge number. A huge number of queries provided by the users can be answered in seconds. It can be done using Natural Language Processing using any language. A custom BLTK tool has to be made using Python. The Naive Bayes classifier and the Edit distance in the Dynamic Programming Method can be employed as algorithms. This may be done using a variety of adapters. Doly, the initial chatbot, is a rudimentary chatbot, capable of little more than introducing herself. Users' affirmations are protected by the

shop whenever they enter them, as are their responses. According to this approach, the closest familiar statement is matched with input, and it then selects an answer from the closest matching familiar reply.[1] In case of invalid answers, the users can press a button and inform the system that the provided answer was invalid. [2]. One of the drawbacks of the chatbot system is that a lot of people are still not much associated with this system, especially those who are not more into technology. In most cases, chatbots are not very accurate in providing the answers or solutions. Sometimes, this system may consume a lot of money. In a few cases, it may lead to a communication gap between the customers and the real service providers of an organization. In most cases, bots can't work without an active internet connection. [2] Conversation success measures can help enhance the performance of chatbots.

Customer chat ratings, average chat lengths, and other Examples of metrics are available. The metrics will be different depending on the purpose of the chatbot and the company's goals, the metrics will be different. These metrics should be compared over time in order to track the bot's performance over time. An overview of the bot's performance will be provided in this analysis. After each chat, chat transcripts can be collected and analyzed. This will enable us to detect any issues that the customers may be going through and where the chatbot is failing. In this way, the bot's weak spots can be identified and optimized to overcome those weak spots. Analyzing a survey based on customer ratings on the bot performance will enable us to get a sense of customer satisfaction. The voice bot may also be taught to overcome its weaknesses. It is possible to design the bot's architecture in a way that makes it appear more human. A machine's capacity to recognize human language words and translate them into a machine-readable format is known as voice recognition. Using a speech recognition technology, we may input data into an application just by speaking into it. Thus, making human-device communication quick and simple. For our proposed system, we will convert the speech of our users to text, then process it using our bot, and finally, produce the desired reply from our bot system. To do this, we can use the polygon smoothing algorithm to process the data, and then use MFCC for feature extraction. In recent times, DNNs (deep neural networks) have become more and more popular in a number of fields, including ASR. So, we can accomplish our task of speech recognition using DNN [3]. DNNs may be used as acoustic models to increase the voice recognition accuracy of a speech recognizer. There are a number of ways to speed up the DNN score computation, allowing for real-time operation. A compressed n-gram language model (LM) can further minimize memory and disk use by doing on-the-fly rescoring [4]. Our chatbot system may also be equipped with Bangla voice recognition. Different Gaussian Mixture Models (GMM) and Deep Neural Network (DNN) based models may be used to build Bangla language audio recognition. With the help of 49 distinct speakers' recordings, a vocabulary of 500 words may be constructed [5].

2.2.2 Literature review on product placement

Generally, store managers want to place the best-selling products on the upper shelves, maybe even in the middle-front row, while products with low profits are shelved at the back or in the lower parts. An investigation undertaken recently [6] by Chen, M., and colleagues (Burke RR Hui Sk & Leykin A) argues the opposite. A

key finding is that consumers' attention is biased in two directions while they are at a store: laterally and vertically. It doesn't contain the order in which things should be arranged on the shelves or how they should be grouped together. As a result, the document will not be thoroughly examined but rather only be mentioned when necessary. The authors found that products that are placed at 14.7 inches below the eye level of customers are the ones that capture their attention the most.

They also argued that customers pay more attention or have a 21% higher propensity to the right side of the aisle while going through it which is applicable to both left and right-handed shoppers. One major drawback of this conclusion is only 11% of the samples used in the report are left-handed. So, there is an overwhelming bias toward right-handed shoppers. This gives us a general idea to prioritize particular shelves while the product placement framework is implemented, as we can then place higher valued products on them based on the output generated by the machine learning model.

Besides, optimal prediction of product placement is a must for the concept to work, and to achieve that, a strongly associated rule mining (ARM) framework is needed, which is discussed by the authors of [7]. They argue that the widely used support-confidence framework of ARM faces major challenges when it is used to generate strong rules. To counter that, an optimized framework is proposed that relies heavily on the "Lift" metric to improve the predictive performance of the associated rule rather than just assuming. One of the major benefits of using this metric is that it helps the seller or manager in determining which products should be placed together. The dataset used here is mainly taken from the transaction of a store, which allowed the authors to identify customer behavior patterns and thus generate outputs that assisted in optimal product placement and identifying products that need promotion.

Aside from the abovementioned Associated rule mining algorithm, strategic product placement can also be done by the PrefixSpan algorithm through mining user buying patterns [8]. The algorithm works in two stages by implementing a pattern-growth framework. Mined product categories are displayed on the shelves in the order they were mined from first-stage data. Second, items from each category are mined and reorganized inside the categories using profit measures. The PrefixSpan algorithm is used to complete the process. From the studies conducted in this paper, it can be seen that there is an increase of average support for l length patterns from 50 to 53 between the two datasets. The same level of output is seen from other length patterns too, which further strengthens the importance of product placement in-store database. However, this framework isn't necessarily the final version. As the author argues, that it can be further optimized by including the concept of revenue optimization to make it more business-friendly. This could be a good addition to our proposed framework as it will help to bridge the gap between the technology & store manager.

As discussed above, dynamic product placement is important for a better customer experience. We can further enhance the framework by understanding the correlation between products in order picking as reported by] Y. Ito and S. Kato, [9]. They argue that travel time for order picking requires a significant amount of time and while various frameworks of product placement have already been studied, there is a lack of it in dynamic situations such as changing of order placement after the order

picking is started & reducing the customer's overall travel time. This also correlates with specific conditions concerning our framework, such as dynamic product placement situations after a customer picks up their order. From the experiments conducted by the authors, there are clear indications of work time being reduced as the existing method needs more travel time than the proposed one. One disadvantage of this experiment is the dataset is small in size & thus may give different outputs if used in a large-scale system. The system is also very basic, and the authors plan to work on it in the future.

C. Bapna et al. [10] proposed an index known as a "generalized utility item set or GUI which extracts generalized high utility or revenues as item sets and then applies it through the proposed framework for product placement. Revenue in retail stores can be improved. They also used taxonomy in their framework. The system evaluation conducted by the authors resulted in an optimized total revenue & execution time. However, a drawback of this proposal may be a lack of design or framework to understand its cost-effectiveness.

Hübner, A et al. [11] argue for a model that optimizes the seller's profit with the help of optimal assortment and also assigns limited shelf space to items. This allows for the methodological arrangement of shelves and increases customer satisfaction. A specialized heuristic is also developed to solve the model. The model is also applicable for both perishable & non perishable products, which is in line with products found in a supermarket. Based on experiments, facings are allocated to either perishable or nonperishable items.

Mining helps us to understand trends, patterns, and correlations, which can help us to build an optimal path for future decisions [12]. This paper basically gives us an in-depth understanding of data mining and gives a general overview of improving associated rule mining (ARM) using modern techniques as well as proposing a multilevel relationship algorithm. The main advantage of this algorithm is that it makes applications that need to run continuously be more efficient. As our proposed framework includes mining to identify product placement, especially the usage of ARM, this paper can help us in understanding and identifying optimal product placement. techniques to be applied in product placement. Suchacka, G. et al.p [13] talk about customer behavior in relation to web server log data. Once again, The document is not going to be examined in-depth and would only be referred to as relevant since we mainly need to understand user purchase likelihood by applying the associated rule. Advantages of this approach include experimenting with both traditional and innovative customers and generating diverse output. It can also adapt to customer behaviors and thus, is very business-friendly. The result clearly shows different motivations for both the groups to generate a high purchase probability. The authors in the future also plan to implement more user session features in the future.

The proposed categorization rules mining method allows for a more thorough examination of incoming data [14]. The study is about designing a data mining approach for analyzing data in a supermarket database to identify where products should be placed on shelves. It introduces a new category of associative rules, which is classification rules. For rules with more than three elements, it enables us to reduce the complexity of the data mining process. As shown in the results, It also allows for a reduction in the number of created dependencies, which benefits the seller. As

such, this approach will not only help us to reduce unnecessary complexities but also enable us to have fewer dependencies, which in turn will result in optimal product placement.

Finally, based on transaction history, a generic framework for predicting customer repeat behavior is proposed by the authors [15]. They propose a metamodel for transactional datasets that abstracts the various data dimensions. This mainly works as a framework for consumer repeat behavior. The results show that the suggestion model “i-prescribe” was able to recommend products with respect to ROI and the limited budget of a particular supermarket. Thus, we can use this strategy to not only make shopping quick for customers but also ensure maximum profit for the seller. Generalized sequential pattern mining known as GSP algorithm is another widely used data mining technique to find sequential patterns from large data. Chengguan and Shihuan have worked on predicting the dynamic cost of enterprise using the GSP algorithm. Here the authors stated that, In the changeable market environment, it is very important that we know how to carry out the informatization to satisfy the consumers’ request for various brands [16]. By doing sequence base pattern mining it is possible to know what customers really want and help the sellers to prepare accordingly. But compared to other data mining methods GSP has some weaknesses. Compared with the AprioriALL Algorithm GSP algorithm counts less candidate sets and the frequent set needn’t be counted in advance in the process of data transformation. Thus GSP Algorithm contributes a lot to the short frequent sequence mode [16]. To tackle this problem the authors have suggested an improved GSP Algorithm which works with QIMS to handle the frequently changing data.

Moreover, the GSP algorithm can be used in different fields even in identifying human behavior. Using videos of different events they have tried to track down patterns in human behavior. A valid option to improve development indicators systems that identify patterns of human behavior is the use of cognitive architectures that allow us to capture some characteristics of human behavior at the level of symbolic and sub-symbolic processing [17]. If video surveillance is combined with the storage capacity, fast access to data, processing capacity and pattern detection, everything is shaken and data is crossed, everything can be done quickly and very fast [17].

2.2.3 Literature review of product sales forecasting

In the customer service management model, sales forecasting is a crucial insight to have. To forecast a time series data, lots of algorithms and hybridization of it can be used. One of which is linear regression, a simple yet famous algorithm for forecasting and shown to provide excellent forecasting with precision rate of 84% and thus proven to work in real world datasets [18].

Another model is random forest and as argued in [19], Random forest is a predicting model that establishes relationship between dependent and independent variable which is highly used for time series finding, predicting and modeling causal effect relation between variables. While working with the data set they have balanced the data set to ensure that there is no biases in the data which does a pretty good job of

predicting nonlinear trends and hence forecasted black friday sales with an average accuracy of 83.6%.

In [20], we see another non linear algorithm which is XG Boost forecasting short term power load. They have stated that it is necessary to do power load forecasting in order to ensure a balance in production and consumption . Here while working with the data set they have kept in mind that the prediction can be affected by a few things . Keeping it in mind they have suggested that the XGBoost algorithm is best for doing this prediction . They have divided the data in four different categories for training and after doing so they have said that the method can divide classification and forecast load value of the appointed date accurately.

2.2.4 Literature review on Smart Contract System

From the paper [21] we can see that ,an increasingly popular use case for smart contracts is the automation of corporate operations in response to events detected by Internet of Things (IoT) sensors, datastreams, or other technologies. Smart contract management systems built on the blockchain are a cutting-edge innovation that can streamline B2B interactions in the near future. It's common knowledge that blockchain plays a crucial role in BPR by enhancing the efficiency of corporate workflow activities, particularly in partnerships with several participants. In this paper, we introduce a method for managing smart contracts that may be used across different businesses. This study is divided into two sections; in the first, we evaluate existing smart contract management platforms based on various criteria that can be crucial when deciding on a solution to meet a certain business need. In the second, we've developed a smart contract and transaction management framework using the blockchain and cutting-edge approaches. A DAO (Decentralized Autonomous Organization) is a type of organization that operates autonomously from its members. This architecture provides a single framework for the execution of smart contracts on the blockchain at the DAO level. The proposed system makes use of two distinct blockchain technologies, namely SBlockchain and TBlockchain. The smart contracts themselves are maintained in the SBlockchain, whereas all the data produced by the smart contracts is kept in the TBlockchain. Moreover, each smart contract has specific terms and stipulations that must be followed in order for the specified event to take place. With the aid of applicable use-cases, we have provided a detailed description of the framework's various components and their implementation.

In the paper [22] it refers to Several blockchain initiatives springing up quickly to combat the COVID-19 pandemic, demonstrating the potential of this technological breakthrough to lessen the impact of the pandemic's multisystemic dangers on emergency management and the economy. This survey delves into the potential applications of blockchain technology in the context of bolstering health measures to curb the spread of COVID-19 diseases and restore normalcy. The survey focuses primarily on the two categories of applications—contact tracing and vaccine/immunity passport support—because these are where blockchains have been used most extensively to offset the effects of COVID-19. The research was conducted with the intention of demonstrating that COVID-19 can only be effectively supported by a mix of blockchain technology and sophisticated cryptographic techniques. Specifically, this

article introduces these methods—zero-knowledge, Diffie Hellman, blind signatures, and proxy re-encryption—before detailing how they are used with blockchains to establish secure and private solutions. Finally, we offer a brief overview of blockchain’s potential uses beyond first contact tracking and vaccination verification.

Chapter 3

Methodology

3.1 System Overview

Here we have discussed about the overall view of our system and for that we have used use case diagram.

3.1.1 System Diagram

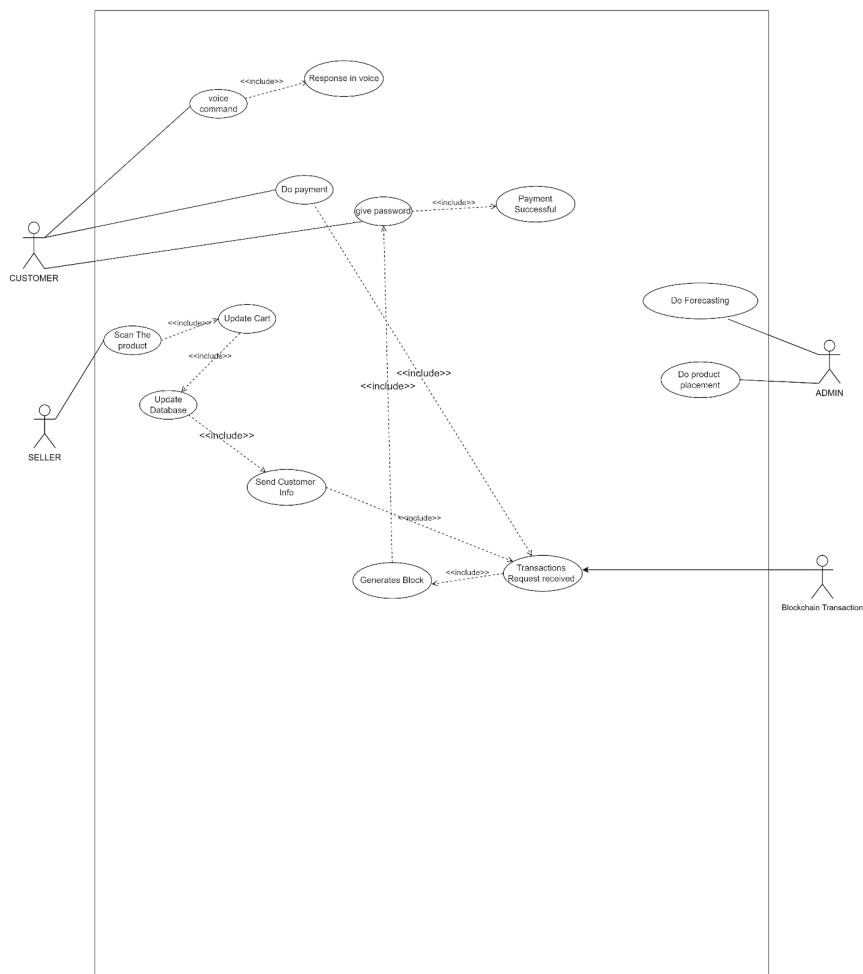


Figure 3.1: Supershop Management System Diagram

3.1.2 System Description

The diagram describes the overall System using HyperAutomation and Machine learning model. There are mainly two primary actors One customer and another is seller. First the customer will interact with the system. Customer will provide the product name as the voice input to the voice bot and the bot will reply the product name and the details of the corresponding product. If the customer uses an inappropriate word it will ask again for the voice input to the customer. The seller will scan the product for selling after that the cart will be updated with the selected product in the database. After that it will generate a request to the Smart Contract System. The Smart Contract system will create a block and it will have all the datas. After that it will ask the user for the Transaction Confirmation notification. Finally The user will authenticate himself and the Transaction will complete and Will have a security.

3.2 Dataset Overview

3.2.1 Data Description

The dataset used here is real world data collected from the PRAN-RFL group/“Daily shopping super shop”. The dataset contains all the information of sales made in the shop from January 2022 to December 2022. The dataset consists of 158293 rows and 9 columns.

As we collected the dataset directly from the company so it was not ready for applying any kind of algorithm. So we have to preprocess the data and get it ready for our use. Which have changed the dimension of our dataset in many cases.

3 Methodologies have been used to do direct implementation of the dataset. These are-

1. Voice Chat Methodology
2. Product Placement methodology
3. Product Sales Forecasting methodology.

While the other 1 methodology has an indirect relationship with the dataset itself. This is-

1. Blockchain transaction system

3.2.2 Data Preprocessing

Product Placement

To get the dataset ready for work we have at first examined if there’s any null value present in our dataset or not. We didn’t find any null values in the dataset but we found a few columns whose values are not useful for us. So we have dropped those columns. Later on we have converted the data of the column that we were using for data mining into a list and applied one hot encoding in our dataset to present our entire dataset in a simpler way.

Product Sales Forecasting (Daily)

First, we did a basic null check and dropped the irrelevant column to ensure a time series dataframe. Then we renamed the “S_CODE” column to “Date” for easier understanding and then converted the date column into date time datatype with Day as the period. Now, as we are interested in the daily product sales we summed up daily total Quantity sold with respect to date column and converted the resulting column into integer type which gave us products sold daily in the supershop with date as the index. Next, we checked the stationarity of the data using augmented Dickey–Fuller test where the resulting p value was less than 0.05. So, we concluded that data is stationary. We also visualized it by doing a time plot.

To implement any of these algorithms individually, we first need to determine the features. For example, lags and time steps. To lag a time series is to move its values one or more time steps forward, or alternatively, to move the times in its index

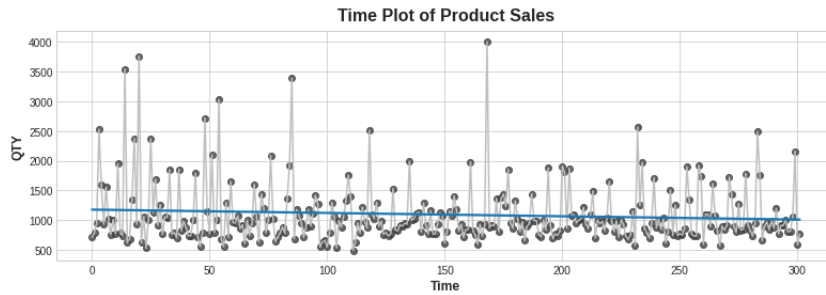


Figure 3.2: Product Sales Time Plotting

one or more steps backward. Time ticks are called TimeSteps. The duration of each samples is what matters.

Now, we used shift function to get the lagged values of daily sales and made a supervised dataset. This dataset then will be applied in the individual algorithms after using scaling techniques.

Hybrid Preprocessing

Here, we first tried to determine the trend/pattern of the dataset by plotting a moving average graph of the entire dataset below:

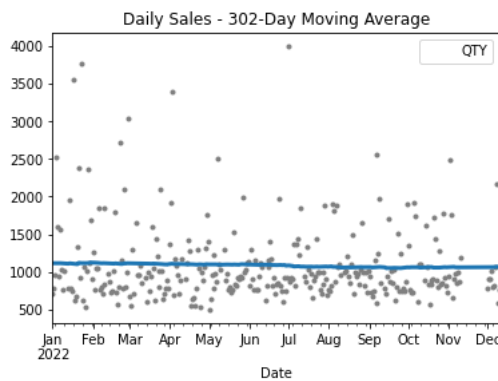


Figure 3.3: Moving Average Graph of the Entire Dataset

Now, this is similar to a linear trend which is stable cause data is stationary. So, we used deterministic process with order = 1. Now the trend generated below is similar to above, so we can conclude that linear trend can be a feature for hybridization technique. Now in both cases of individual and hybrid implementation, we used to test 90 days and rest of the day before $(301-90) = 211$ days were used as training data.

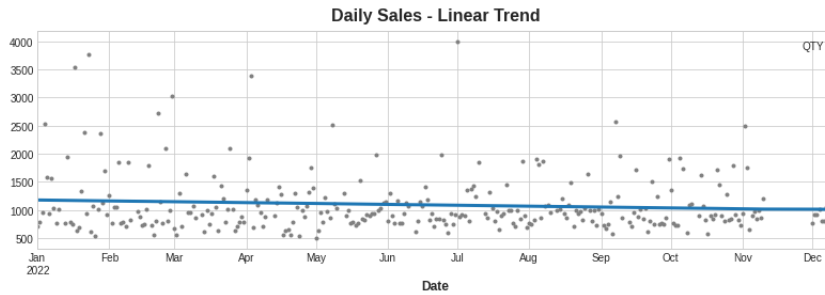


Figure 3.4: Linear Trend Generation

3.3 Model

3.3.1 Voice Chat Bot Methodology and Implementation

Since Voice chat Bot is a well-known AI based software used by many successful software companies such as Apple’s use Siri, Google has Google Assistant, Amazon uses Alexa etc so therefore we decided to add voice chat bot in this hyper-automation project where our Voice chat bot will primarily work/function based on the 3 key concepts to engage in conversation with our customers and respond their queries about super store. These are Voice Transformation, Natural Language Processing and Multi-Language Translation.



Figure 3.5: The concepts of Voice Chat Bot

Voice transformation

Voicebots respond to spoken instructions or requests. After getting the input voice, the voicebot system tries to recognize the voice and then transfer it into text in order to check the queries. The voicebot system then responds to the queries by text and finally converts those responses into voice. That is how a voicebot’s standard speech dialogue operates.

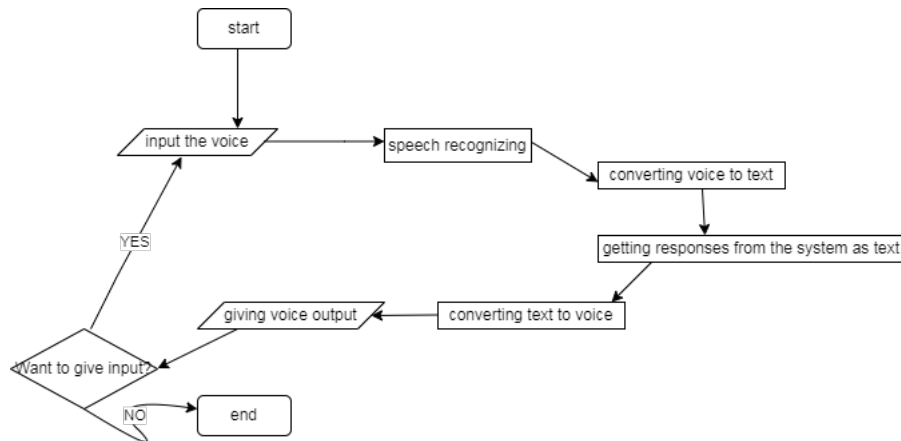


Figure 3.6: Flowchart Diagram of a general Voice Transformation

Natural Language Processing

The capacity of a computer system to comprehend and interpret human language is known as natural language processing (NLP). However, human language is frequently complicated, confusing, and full of slang and bad grammar. In order to be able to deduce a user's intents from written or spoken language even when it is imperfect, voicebot systems need NLP.

The most well-known Python Natural Language Processing tools are NLTK and spaCy. NLTK is a library for string manipulation. It accepts strings as input and outputs either strings or lists of strings, while spaCy employs an object-oriented methodology. The document object that spaCy produces after text parsing contains words and phrases that are objects in and of itself. The performance of spaCy is superior in word tokenization and POS-tagging, whereas NLTK surpasses spaCy in sentence tokenization. The optimal technique to do a task is provided by spaCy, despite the fact that NLTK offers access to many other algorithms.

Multi Language Translation

A free and limitless Python package called googletrans uses the Google Translate API. Google offers a machine translation service called Google Translate API. It enables programmers to include automated translation features in their tools, websites, and apps. With the help of the API, programmers may identify the language of a piece of text and translate it from one language to another. The API leverages Google's neural machine translation technology, which is made to provide accurate translations, and it covers a large number of languages. Additionally, it has a function called "auto-detection" that determines the text's original language without the developer having to define it.

WorkFlow

In order to create an optimal voice chat bot, several libraries must be installed. First, PyAudio Library must be installed in order for the software system to obtain a microphone in order to receive and transmit speech. The audio must then be converted to text. In order to achieve that, we installed the SpeechRecognition Library, a branch of the Natural Language Processing area (NLP). Third, we set up

the pandas libraries. We established a dictionary with the assistance of the pandas library and the ML command that will save all the data from a csv file so that we can utilize it for further purposes. The fourth step in creating a voice chatbot is to install the spacy library. Spacy, which follows an object-oriented approach, is superior in terms of word tokenization and POS tagging when compared to NLTK's string processing approach. By utilizing the POS tagging feature of the spacy library, we traverse the input text (which was originally in voice form but was then converted to text) word by word and in our code "word.pos_" returns what type of word it is (such as a noun, pronoun, verb, punctuation, etc.) which helps us to create if-else conditions to extract necessary information (such as any specific strings, product names, and category names) from the input text and store them in variables. The fifth step in creating a voice chatbot is to install the NLTK library and use its 'bleu' function to check the similarities of the necessary information previously that saved on the variables and the informations like product name, category name, product price etc which was saved in the dictionary by traversing that dictionary. When the voicebot system has finished traversing the dictionary, it will deliver a text answer based on which if elif else condition yields the most similarities. The voicebot will then convert the English text into Bangla text using the Google Translate API, which was accessed via the googletrans library. The voicebot will save the English and Bengali text responses in variables. Finally, the pyttsx3 library is installed to give the voicebot the ability to convert text to speech. As a result, the voicebot will convert both Bengali and English text into speech and deliver them one by one. That's the overview of code and workflow of how our voicebot is working.

3.3.2 Product Placement Methodology and Implementation

Market basket analysis which is considered as data mining approach that can be utilized in any retail environment to find buying trends. Mathematical affinity calculations enable company owners to better understand their clients and ultimately serve them by emphasizing their purchase habits. To put it simply, MBAs are interested in what items are most usually purchased together. You may leverage these relationships for cross-selling, suggestions, promotions, or indeed the positioning of products on a menu or in the shop to maximize profitability.

The algorithm used by us are:

1. Apriori algorithm
2. FP Growth Algorithm
3. GSP algorithm

Implementation

Apriori, FP Growth, GSP algorithm are known as market basket algorithms to determine sequence patterns from any random dataset. We have used all three algorithms in our dataset which contains the daily sales information of a super shop. By going through the data of first ten months sales we have tried to find out the relations between different products. Based on those relations it will be easy for any shop to choose which products should be kept with which product. For example, if {A}, {B} and {C} are three products we will find out relations between these products and based on those relations we will make our decision on placing them. If

the algorithms are showing results which show $\{A>C\}$ have better relations among them compared to $\{A>B\},\{B>C\}$ we will place product $\{A\}$ and $\{C\}$ together to make it easier for the customer to find them easily as most of them buy these products together.

When working with a particular set of transaction data association rule mining algorithms can easily find out the rules which can help the shopkeepers to determine the sale of a particular product based on the other item in the transaction. To extract these rules data mining algorithms use a few metrics like Support , Confidence and lift.

Support refers to the frequency of an item set in the transaction dataset.

$$Support\ of(A \rightarrow B) = \frac{(A \cup B)}{n}$$

Confidence indicates how often a rule is appearing to be true .

$$Confidence(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A)}$$

Lift is a measure that tells us if the probability of consequent increases or decreases given the purchase of antecedent .

$$Lift\ of(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A) \times Support(B)}$$

We set minimum threshold values for these metrics if a rule's metrics exceeds the threshold limit we accept that rule. Setting different thresholds minimum support and minimum confidence values we have checked which algorithm is finding all the rules in less time compared to other algorithms.

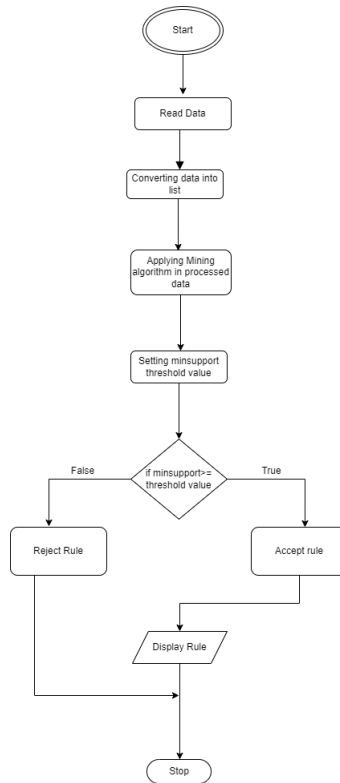


Figure 3.7: Workflow Diagram of Product placement Methodology

3.3.3 Product Sales Forecasting and Implementation

Forecasting of time series data is one of the most important insights to have in the business world. It helps businesses to understand the needs of the customers and react accordingly. In this research 3 algorithms chosen to do forecasting are listed below:

Linear regression

Linear regression is a highly used algorithm in statistics it is used in describing relations between a scalar variable and one or more explanatory variables. Simple linear regression deals with a scalar variable and one explanatory variable . The regression model does the work of finding if the predictor variable is predicting the outcome properly or not . The simple linear regression can be defined by the formula $Y=bx+c$. Here, y defines the output dependent variable score, b means regression coefficient, c means constant and x is the independent variable.

Regression models are one of the most used algorithms for prediction . The algorithm has been widely used for weather forecasting and trend forecasting.

Random forest

Random forest is another algorithm used by us which is popular in machine learning for solving Regression and Classification problems. Random forest works with a number of decision trees to tackle complex problems . It takes n number of decision

trees based on the dataset and provides a combined average value which helps to generate an improved prediction model.

XGBOOST

Xgboost algorithm is an open source implementation of gradient boost tree algorithm. It provides us with a scalable ,portable and Distributed gradient boosting library. The algorithm adapts very well with all kinds of data and it helps us to predict any sort of data quite easily . The algorithm is a combination of multiple ml algorithms which provides the user a better result. Random forest and XG Boost algorithms are quite the same as both work with multiple layers of decision trees . The difference can be seen in the structure of the trees. In the XG Boost algorithm the trees are built in a parallel manner.

Implementation

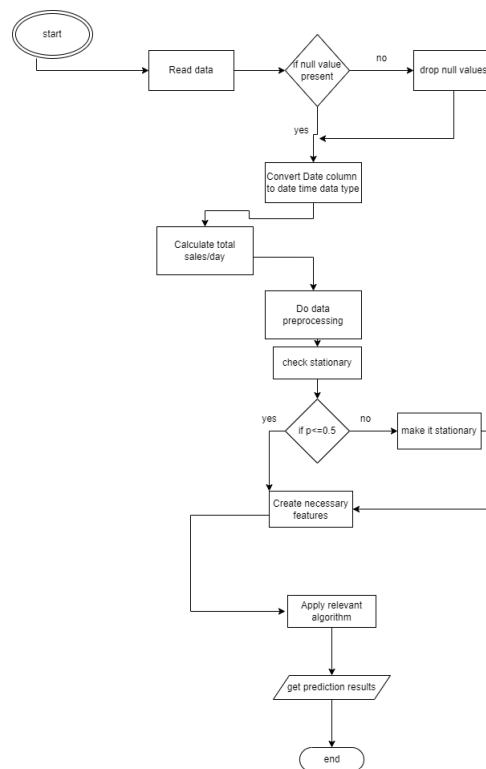


Figure 3.8: Workflow of Forecasting

After importing the libraries and doing necessary preprocessing we first implement the linear regression algorithm. As we already discussed above, the simple linear regression can be written as $y = (\text{weights} * \text{features}) + \text{bias}$, where for each feature, the algorithm learns weight and based on the best fit target it selects weight and bias during training. So, after training, we fit the model and then predict it which give us an approximate forecast to test with test dataset.

Next, we move on to random forest which is a tree based algorithm based on decision tree. Although this is a famous classification model, we found that this model worked very well with our dataset and that can be attributed to different individual

trees output being averaged. And the training here is done by bagging method. This also ensures that overfitting is few. Now, this algorithm chooses it's root node based upon gini index. Through this we can understand the impurity of dataset. The

$$\text{formula for which is } 1 - \sum_{i=1}^n (P_i)^2$$

Then we calculate the weighted gini index which is the total gini index of a particular branch. The feature with lowest gini index will be chosen as the root node.

Entropy is another way to calculate the impurity. Now, we just normally apply the model after setting up n_estimators and max depth parameter where the first one is number of decision trees to be used and the second is the depth of such trees. More, trees generally give good outputs but take more time.

Next one is XGboost which is a gradient boosted decision tree algorithm. Thus, this also depends upon decision trees. Unlike, random forest, XGboost takes leaf node of a tree and may turn that node into a if condition if it benefits the model which is determined by the loss function. This algorithm also uses backward pruning after using max depth parameter. The loss function is as follows

$$\sum_{i=1}^n l(y_i + \hat{y}_i ft(x_i)) + \Omega(ft)$$

Hence, enhances the overall performance

After that we did hybridization of linear-random forest and linear-xgboost to see whether the performance improves. In both cases we fit linear regression with original series and the 2nd algorithm to the residual series. Because linear regression knows how to extrapolate the trend while the tree based algorithms are good at learning interactions

The process is as below:

First, we train and predict with linear regression. Then we predict the X train-1 on linear regression. Later, we fit the 2nd model on the residual series which is:

Train model of target series - predicted series from the first model. Then with this, we fit the 2nd model with a new feature (X Train-2) values and predict that. Now, we add both prediction to get the overall prediction.

So, here, as stated above we got linear trend, so using linear regression, we train it on the trend and the overall trend for out of bound sample is:

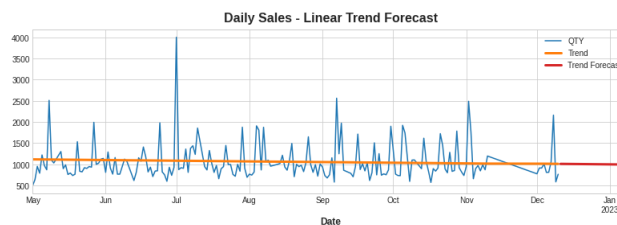


Figure 3.9: Linear Trend forecast

So, here we extrapolate the trend and then remove it by transforming y and apply the next model on the error series.

3.3.4 BLOCKCHAIN: Smart Contract System process and Implementation

Required Tools

For the Simulation environment, Ganache was used with Truffle framework to assest pipeline for the blockchain. Solidity was used for coding.

WorkProcess

At first the main system have to connect with the blockchain. After that the system will receive the Customer ID and the shopping list as Input and Generate a new block which will contain all the necessary information of the Smart Contract. Moreover, Customer count is a Global variable and after that One customer data will be stored and the count will increase. We are going to execute a migration script before putting the smart contract onto the blockchain to make it live. We are unable to add new contracts to the blockchain if we do not have it since the blockchain is impervious to any changes. As our Whole System will be controlled by blockchain So we need another Class for it. As a first step, we'll verify that the smart contract has been successfully deployed and that we can access the tasks using the task count. Firstly we have to verify that the contract's address is not an empty or undefined part of the deployment check. Then, verifying the listing of tasks by comparing the product is valid, scannable, and available in the database or not.

Here, after putting the newly created task through its paces to see if it works. An external client might choose to subscribe to the events generated by the creation of a new task. They may build tests to examine the information and guarantee that the events are triggered on every new activity. Here event ID will increase by $n+1$. As now the new Customer Creation is done finally we need a payment complete confirmation. Finally, while doing payment the system will send a pop notification to the customer to clear the payment. After that While sending the money from the mobile banking system, the user will input his password to verify himself.If the password is matched then it will show the payment complete notification. Whereas if any info is wrong then the system will not clear the payment, redirect to the previous page and count it as unpaid customer. All the necessary information will be stored in the block. Finally, According to Smart contract principle As the One side product buying is done it will cut off the necessary payment from the user account and the Transaction will be secured as there is no intermediary's involvement for the transactions.

WorkFlow Diagram

Algorithm 1 Pseudo Code for Generating new customer and shopping product List

Connect \rightarrow (blockchain)
customerCount = 0

Input:

Initialize customer_ID
Initialize shopping_list

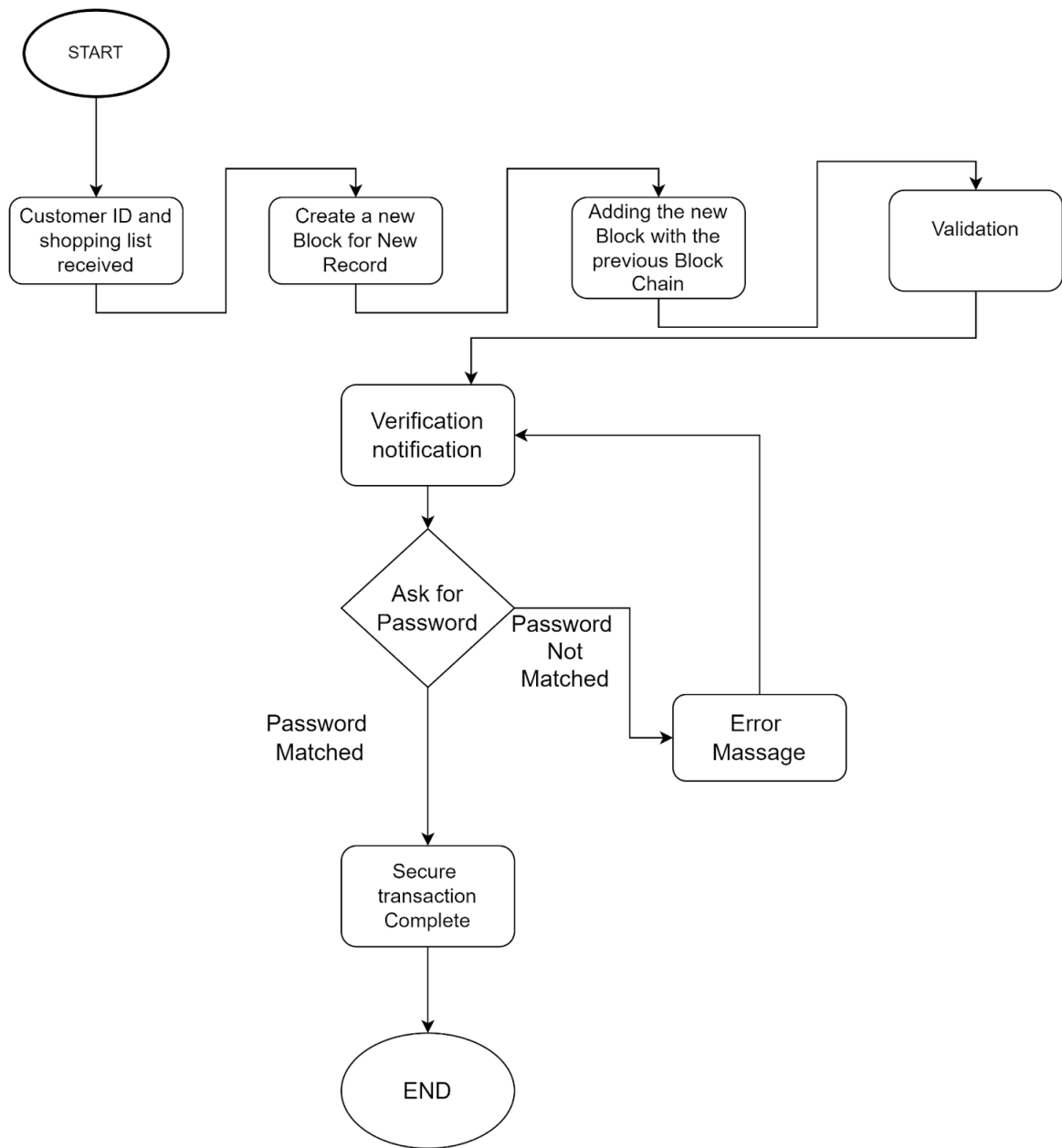


Figure 3.10: Workflow Diagram of Smart Contract System Transactions Method

bool_completed

Output:

```
customerCount + +  
shopping_list  
False  
End();
```

Algorithm 2 Pseudo Code for testing 1 Cart listing product and deployment into Blockchain

Connect→(blockchain)

Connect→(customer accounts)

if deployment == is successful then

check if:

```
payment_information is not null;  
payment_information is not 0;  
payment_information is not undefined;
```

if list_products == successful then

check if:

```
id of the product is valid;  
product is scannable;  
product data is available in database;
```

End();

Algorithm 3 Pseudo Code for testing 2 new customer creation

Connect→(blockchain)

Connect→(customer accounts)

if new_customer_creation == successful then

```
create new product_list;  
check customer count for new customer;
```

check if:

```
event is getting triggered;  
event id is n+1;  
event completion is false;
```

End();

Algorithm 4 Pseudo Code for testing 2 (payment completion)

Connect→(blockchain)

if bool_completed == true **then**

 payment is completed;

 payment completion function is true;

 check if:

 event is getting triggered;

 authenticate password;

 first customer has completed payment and event id is 1;

 event completion is true;

End();

Chapter 4

Result Analysis

Result Analysis

Voice Chat Bot

Our voicebot was developed using the ideas of speech transformation, natural language processing, and multilingual translation.

Our voice bot uses the speech recognition library and the pyttsx3 library to convert speech to text and text to speech, and it does it flawlessly.

To implement NLP in our voicebot we used spacy library and nltk library,for word tokenization we used spacy library instead of nltk library.For tokenization and other typical NLP tasks, spaCy is often quicker than NLTK in terms of time complexity. This is so because Cython, a computer language that is a superset of Python and may be translated to C for better performance, is used to create spaCy. Additionally, pre-trained models that are speed-optimized are used by spaCy, which boosts performance even more.In contrast, NLTK is created entirely in Python, which might be slower than Cython. Additionally, NLTK does not come with pre-trained models; as a result, you will have to spend more time training models yourself.spacy is faster than NLTK for tokenization, part-of-speech tagging, and named entity recognition. spaCy also includes a built-in visualizer called displaCy that allows you to visualize the output of its NLP models. If performance and ease of use are your top priorities, spaCy is a better choice.That's why for word tokenization we used poss tagging concept for word tokenization.Moreover to detect word similarity we used nltk's bleu library which gives 100% accuracy to find the similarities between 2 strings.

And lastly for language translation we used googletrans library to implement Google translator API which successfully convert one language to another language to deliver multi-language voice in our Voice Chat Bot.

Product placement

Fp growth algorithm is performing better compared to Apriori and GSP algorithm. We can see that Apriori and Fp growth algorithms are taking almost the same time

Algorithm	Minimum Support	Time
Apriori	0.03	1.126 s
	0.05	0.413 s
FP growth	0.03	0.473 s
	0.02	0.493 s
GSP	0.03	26.418
	0.02	86.762

Table 4.1: Product Category

Algorithm	Minimum Support	Time
Apriori	0.002	3.493 s
	0.005	0.428 s
FP growth	0.002	2.316 s
	0.005	1.927 s
GSP	0.002	10530.413
	0.005	4255.696

Table 4.2: Product Name

to produce the rules while checking with different minimum support values but Fp Growth is producing a comparatively higher number of rules. While using Apriori algorithm the runtime increases exponentially depending on the number of transactions. But the FP Growth algorithm increases its runtime linearly depending on the number of transactions. Another reason for the FP Growth algorithm performing better is it iterates through the dataset only twice to generate the rules but the gsp and apriori algorithm iterates through the dataset multiple times which consumes more time.

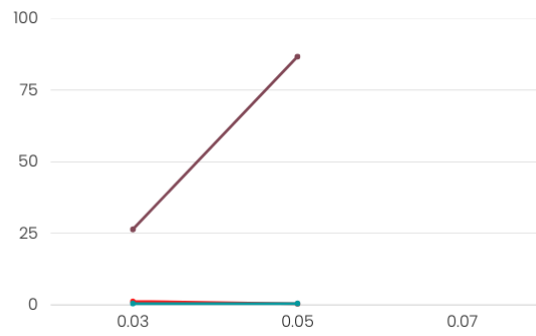


Figure 4.1: product category performance

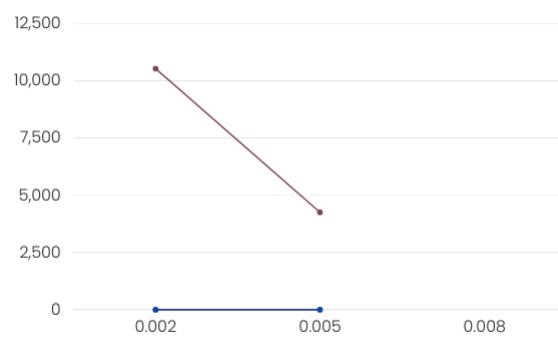


Figure 4.2: product name performance

As FP growth algorithm is performing better in comparison to other two we are suggesting to use this algorithm for finding product placement rules.

Product Sales Forecasting

We have used mean absolute error, root mean squared error, mean absolute percentage error and lastly r-squared to measure the performance of the algorithms. Interestingly, in all 3 individual cases we saw below average performance. As discussed above, this is due to linear regression not being good at interacting with the data while the tree algorithms are unable to extrapolate the trend.

But, when we implemented hybridization, both models gave us 90 percent plus accuracy. This may be the case of XGBoost and random forest not knowing how to extrapolate a trend beyond the training set and thus linear regression helps those two algorithms learn about it.

Mean absolute error evaluates the distance between forecasted value and actual value. But there is a chance of getting negative results here. We can also use mean squared error to evaluate the distance but here the unit gets squared. So to remove that we use root mean squared error. And lastly, we used mean absolute percentage error to get the overall error to get a proper understanding of the forecast. But here MAPE did not give us clear answer as to which hybrid model is performing better. So, we used r squared for comparison. In a regression model, R-Squared (also known as R2 or the coefficient of determination) is a statistical metric that quantifies the percentage of the dependent variable's variation that can be accounted for by the independent variable.

$$MAPE = 1 \div n(\sum_{i=1}^n |(At - Ft) \div At|)$$

Where At is actual value and Ft is forecasted value and n is number of time for summation iteration

$$MAE = (\sum_{i=1}^n |yi - xi|) \div n$$

$$RMSE = \sqrt{\sum_{i=1}^n (xi - \bar{xi})^2} \div n$$

$$R - Squared = 1 - (RSS \div TSS)$$

where xi is observations from time series and xi bar is the estimated time series while n is number of non missing data point.

Now, let's check the performance of the models using these metrics

Algorithm	MAE	RMSE	MAPE
Linear Regression	343.52	427.80	0.33
Random Forest	363.71	513.90	0.34
XGBoost	346.66	498.37	0.32
Linear Regression-Random Forest	82.71	93.58	0.077
Linear Regression-XGBoost	68.94	77.48	0.09

Table 4.3: Performance of Different Machine learning Models

The first 3 algorithms show result individually like the figure below and the last 2 figure represent the hybrid models. As seen, the hybrid linear regression-XGBoost is outperforming every other algorithm if we take the value of R-Squared into account which are 0.963 for linear-xgboost regression and 0.943 for the other hybrid model. Hence, we propose using this as the basis for any custom model and to complete current forecasting.

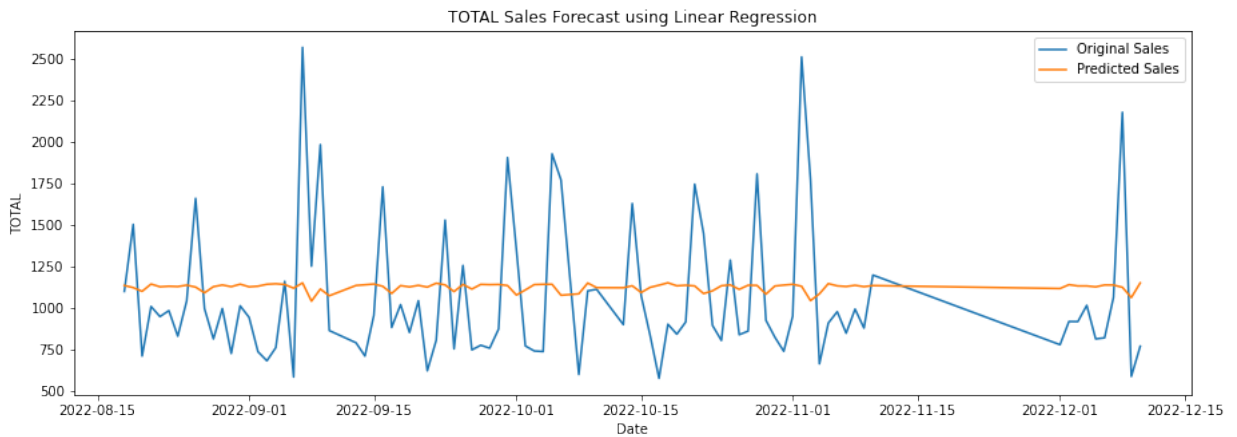


Figure 4.3: Forecasting using linear regression

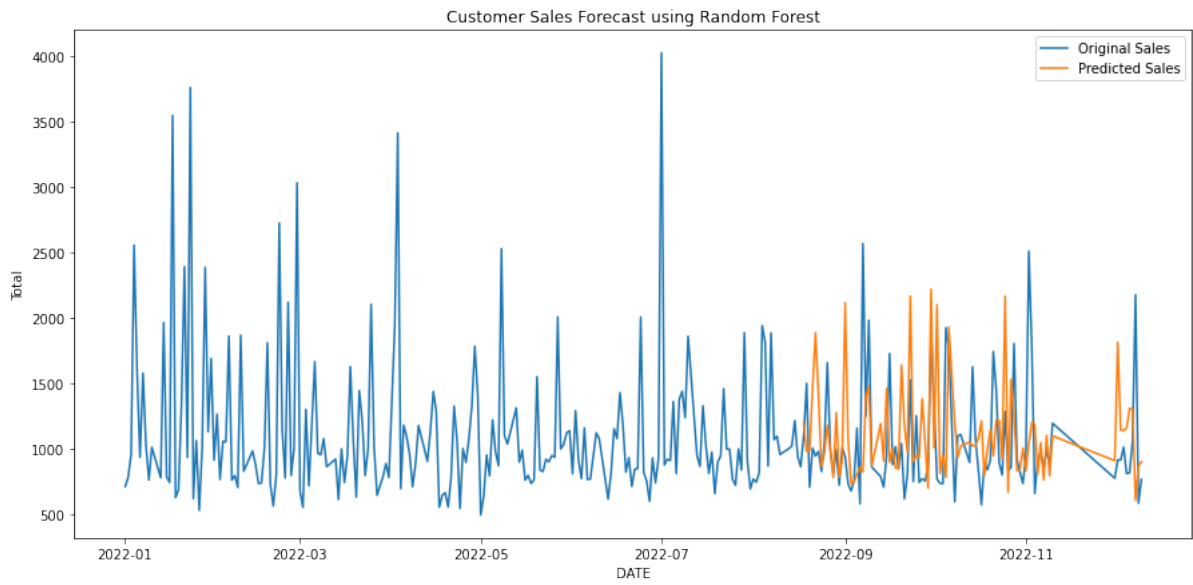


Figure 4.4: Forecasting using random forest

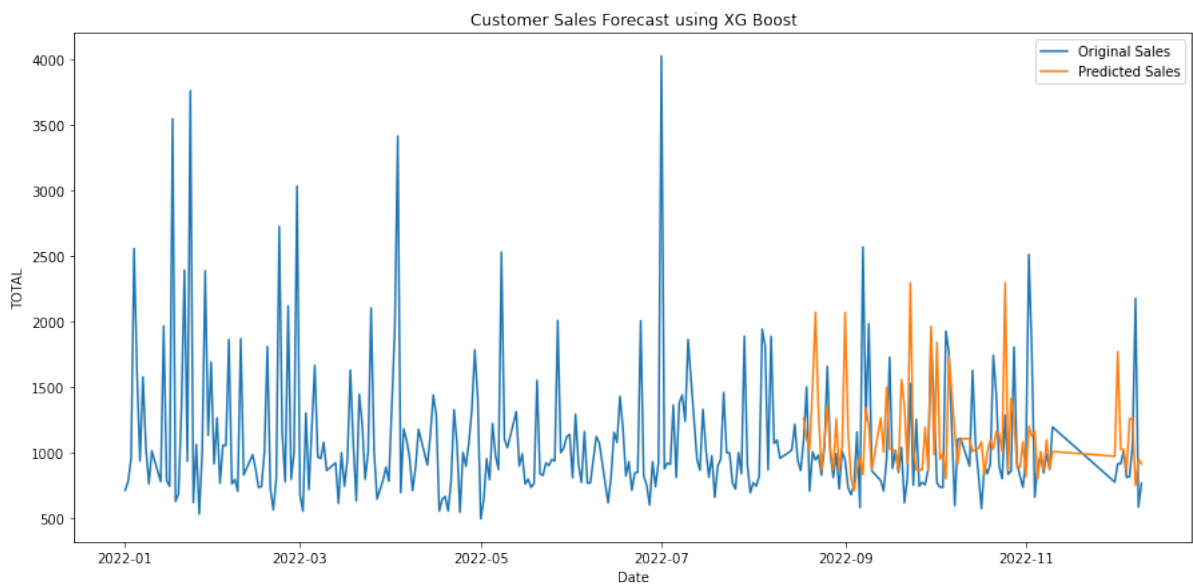


Figure 4.5: Forecasting using xgboost

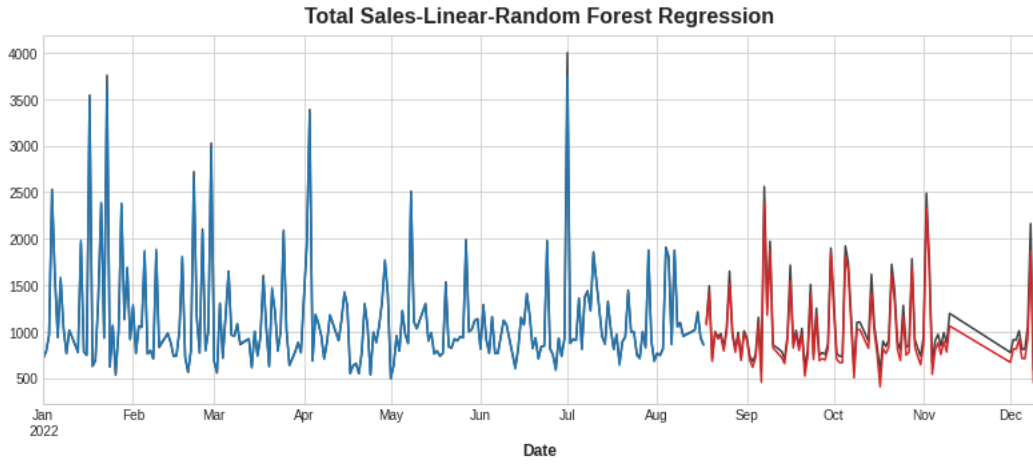


Figure 4.6: Forecasting using linear regression-random forest

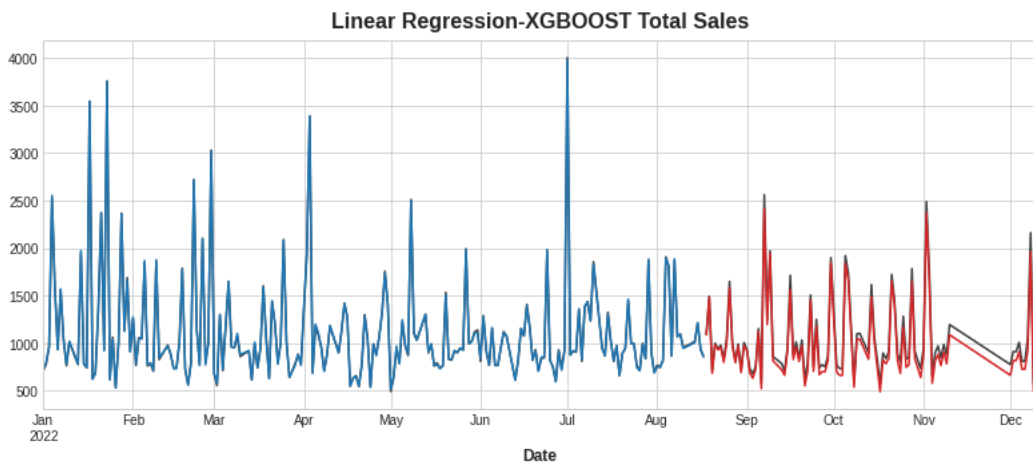


Figure 4.7: Forecasting using linear regression-xgboost

BLOCKCHAIN: Smart Contact System

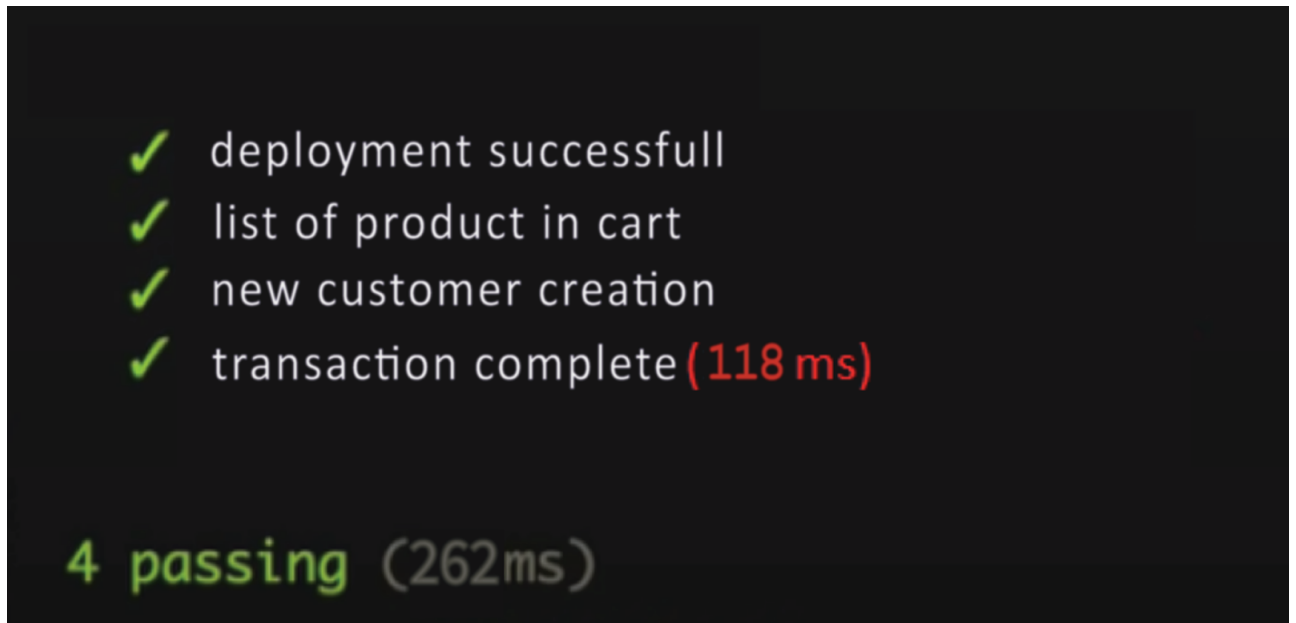


Figure 4.8: Successful Transaction

Here we can see our algorithms has successfully executed so The System has satisfied the Smart Contract System requirement. As a result, the secure Transaction has been executed.

Chapter 5

Conclusion

Conclusion

In this modern era, everything is getting updated by the touch of technology, so why not the supermarket? To make life easier, we go to the super shop to buy our daily needs. So to make the super shop much more useful, easy to explore, and customer friendly we decided to implement hyper-automation in our super shop system. In our thesis project, we will use different kinds of algorithms so that the customer can feel comfortable easily using our super shop system, moreover, we will also try to improve the response time of the system so that the system can respond to the customer quickly and more accurately. Furthermore, based on the customer's priority the system will suggest the product to the customer so that customer may get interested to buy that product and so on. Although we have already introduced you to so many new ideas that we are going to implement in our super shop system, we think it's just the beginning. We need to work on more data sets, ideas, and more algorithms to advance the system of our superstore. In the future, we hope to have a better version of the system by adding more useful features with high efficiency.

5.0.1 Future Work

Due to the limitation of time we can not implement all the things . So one of the major things is to use the Crowd Management System in a counter. So counter management system is our first priority. Moreover, we will use explainable AI framework to explain the performance of the models. Hence our system is decentralized so we need a proper centralized system for the super shop. For the Bot part we would like to add rating and review system in the voice bot. Also, we would like to implement a custom model based on linear-XG Boost regression for more precision forecasting and further analyze different types of features related to business such as profit. Moreover, we would also like to make a custom model for product placement and try to do real life testing of it in a super shop and compare new data to see if profit/sales increased. This would also require a large new data set so that both forecasting and product placement models can be trained on big real world data. Lastly, as there is no checking system to check if the currency is fake or not, we would like to implement the Machine learning models for detecting the fake notes.

Bibliography

- [1] M. Kowsher, F. S. Tithi, M. Ashraful Alam, M. N. Huda, M. Md Moheuddin, and M. G. Rosul, “Doly: Bengali chatbot for bengali education,” in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh: IEEE, May 2019, pp. 1–6, ISBN: 9781728134451. DOI: 10.1109/ICASERT.2019.8934592. [Online]. Available: <https://ieeexplore.ieee.org/document/8934592/>.
- [2] A. Tiwari, R. Talekar, and S. Patil, “College information chat bot system,” *International Journal of Engineering Research and General Science*, vol. 2, no. 2, 2017.
- [3] S. Virkar, A. Kadam, N. Raut, S. T. Shohaib Mallick, and P. Pimpri Chinchwad College of Engineering, “Proposed model of speech recognition using mfcc and dnn,” en, *International Journal of Engineering Research and*, vol. V9, no. 05, p. 05, May 2020, ISSN: 2278-0181. DOI: 10.17577/IJERTV9IS050421. [Online]. Available: <https://www.ijert.org/proposed-model-of-speech-recognition-using-mfcc-and-dnn>.
- [4] X. Lei, A. Senior, A. Gruenstein, and J. Sorensen, “Accurate and compact large vocabulary speech recognition on mobile devices,” en, in *Interspeech 2013*, ISCA, Aug. 2013, pp. 662–665. DOI: 10.21437/Interspeech.2013-189. [Online]. Available: https://www.isca-speech.org/archive/interspeech_2013/lei13_interspeech.html.
- [5] J. Rahman Saurav, S. Amin, S. Kibria, and M. Shahidur Rahman, “Bangla speech recognition for voice search,” in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, Sylhet: IEEE, Sep. 2018, pp. 1–4, ISBN: 9781538682074. DOI: 10.1109/ICBSLP.2018.8554944. [Online]. Available: <https://ieeexplore.ieee.org/document/8554944/>.
- [6] M. Chen, R. R. Burke, S. K. Hui, and A. Leykin, “Understanding lateral and vertical biases in consumer attention: An in-store ambulatory eye-tracking study,” en, *Journal of Marketing Research*, vol. 58, no. 6, pp. 1120–1141, Dec. 2021, ISSN: 0022-2437, 1547-7193. DOI: 10.1177/0022243721998375. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/0022243721998375>.
- [7] O. Olasehinde Olayemi, A. Abiona Akeem, and A. Ibiyomi Michael, “Design of machine learning framework for products placement strategy in grocery store,” *Communications*, vol. 7, pp. 5–11,
- [8] G. Aloysius and D. Binu, “An approach to products placement in supermarkets using prefixspan algorithm,” en, *Journal of King Saud University - Computer and Information Sciences*, vol. 25, no. 1, pp. 77–87, Jan. 2013, ISSN: 13191578.

DOI: 10.1016/j.jksuci.2012.07.001. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1319157812000353>.

- [9] Y. Ito and S. Kato, “Dynamic product placement method in order picking using correlation between products,” in *2016 IEEE 5th Global Conference on Consumer Electronics*, Kyoto, Japan: IEEE, Oct. 2016, pp. 1–3, ISBN: 9781509023332. DOI: 10.1109/GCCE.2016.7800361. [Online]. Available: <http://ieeexplore.ieee.org/document/7800361/>.
- [10] C. Bapna, P. K. Reddy, and A. Mondal, “Improving product placement in retail with generalized high-utility itemsets,” in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, Sydney, Australia: IEEE, Oct. 2020, pp. 60–69, ISBN: 9781728182063. DOI: 10.1109/DSAA49011.2020.00018. [Online]. Available: <https://ieeexplore.ieee.org/document/9260086/>.
- [11] A. Hübner and K. Schaal, “An integrated assortment and shelf-space optimization model with demand substitution and space-elasticity effects,” in *European Journal of Operational Research*, vol. 261, no. 1, pp. 302–316, Aug. 2017, ISSN: 03772217. DOI: 10.1016/j.ejor.2017.01.039. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0377221717300747>.
- [12] D. Vidhate and P. Kulkarni, “To improve association rule mining using new technique: Multilevel relationship algorithm towards cooperative learning,” in *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, Mumbai, Maharashtra, India: IEEE, Apr. 2014, pp. 241–246, ISBN: 9781479924943. DOI: 10.1109/CSCITA.2014.6839266. [Online]. Available: <http://ieeexplore.ieee.org/document/6839266/>.
- [13] G. Suchacka and G. Chodak, “Using association rules to assess purchase probability in online stores,” in *Information Systems and e-Business Management*, vol. 15, no. 3, pp. 751–780, Aug. 2017, ISSN: 1617-9846, 1617-9854. DOI: 10.1007/s10257-016-0329-4. [Online]. Available: <http://link.springer.com/10.1007/s10257-016-0329-4>.
- [14] P. Pukach and K. Shakhovska, “The mathematical method development of decisions supporting concerning products placement based on analysis of market basket content,” in *2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, Lviv - Polyana, Ukraine: IEEE, 2017, pp. 347–350, ISBN: 9781509050451. DOI: 10.1109/CADSM.2017.7916147. [Online]. Available: <http://ieeexplore.ieee.org/document/7916147/>.
- [15] A. H. Kazmi, G. Shroff, and P. Agarwal, “Generic framework to predict repeat behavior of customers using their transaction history,” in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, IEEE, 2016, pp. 449–452.
- [16] C. Xiang and S. Xiong, “The gsp algorithm in dynamic cost prediction of enterprise,” in *2011 Seventh International Conference on Natural Computation*, IEEE, vol. 4, 2011, pp. 2309–2312.

- [17] H. F. Gomez A, V. H. Cordova A, E. F. Robalino P, *et al.*, “Application of the gsp_m algorithm for the identification of behavioral patterns of people who shoplift,” in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Huangshan, China: IEEE, Jul. 2018, pp. 1036–1044, ISBN: 9781538680971. DOI: 10.1109/FSKD.2018.8687200. [Online]. Available: <https://ieeexplore.ieee.org/document/8687200/>.
- [18] G. T., R. Choudhary, and S. Prasad, “Prediction of sales value in online shopping using linear regression,” in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India: IEEE, Dec. 2018, pp. 1–6, ISBN: 9781538669471. DOI: 10.1109/CCAA.2018.8777620. [Online]. Available: <https://ieeexplore.ieee.org/document/8777620/>.
- [19] H. V. Ramachandra, G. Balaraju, A. Rajashekar, and H. Patil, “Machine learning application for black friday sales prediction framework,” in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India: IEEE, Mar. 2021, pp. 57–61, ISBN: 9781728185194. DOI: 10.1109/ESCI50559.2021.9396994. [Online]. Available: <https://ieeexplore.ieee.org/document/9396994/>.
- [20] Y. Liu, H. Luo, B. Zhao, X. Zhao, and Z. Han, “Short-term power load forecasting based on clustering and xgboost method,” in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China: IEEE, Nov. 2018, pp. 536–539, ISBN: 9781538665657. DOI: 10.1109/ICSESS.2018.8663907. [Online]. Available: <https://ieeexplore.ieee.org/document/8663907/>.
- [21] M. Muneeb, Z. Raza, I. U. Haq, and O. Shafiq, “Smartcon: A blockchain-based framework for smart contracts and transaction management,” *IEEE Access*, vol. 10, pp. 23 687–23 699, 2022, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3135562. [Online]. Available: <https://ieeexplore.ieee.org/document/9650875/>.
- [22] L. Ricci, D. D. F. Maesa, A. Favenza, and E. Ferro, “Blockchains for covid-19 contact tracing and vaccine support: A systematic review,” *IEEE Access*, vol. 9, pp. 37 936–37 950, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3063152. [Online]. Available: <https://ieeexplore.ieee.org/document/9366740/>.