KANTIPUR ENGINEERING COLLEGE

(Affiliated to Tribhuvan University)

Dhapakhel, Lalitpur

[Subject Code: CT654]

A MAJOR PROJECT PROPOSAL ON

END-TO-END ENCRYPTION AND DECRYPTION USING CRYPTOGRAPHY (ASYMMETRIC KEY ENCRYPTION AND DECRYPTION (RSA))

Submitted by:

Anup chaudhary [KAN075BCT009]

Chris Gurung [KAN075BCT023]

Himalaya Pal [KAN075BCT029]

Kundan Giri [KAN075BCT030]

A MAJOR PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR IN COMPUTER ENGINEERING

Submitted to:

Department of Computer and Electronics Engineering

June, 2022

END-TO-END ENCRYPTION AND DECRYPTION USING CRYPTOGRAPHY (ASYMMETRIC KEY ENCRYPTION AND DECRYPTION (RSA))

Submitted by:

Anup chaudhary [KAN075BCT009]

Chris Gurung [KAN075BCT023]

Himalaya Pal [KAN075BCT029]

Kundan Giri [KAN075BCT030]

A MAJOR PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR IN COMPUTER ENGINEERING

Submitted to:

Department of Computer and Electronics Engineering

Kantipur Engineering College

Dhapakhel, Lalitpur

June, 2022

COPYRIGHT

KANTIPUR ENGINEERING COLLEGE
DEPARTMENT OF COMPUTER AND ELECTRONICS ENGINEERING

APPROVAL LETTER

The undersigned certify that they have read and recommended to the Institute of Engineering for acceptance, a project report entitled ”End-To-End Encryption and Decryption Using Cryptography (Asymmetric Key Encryption and Decryption (RSA)) ” submitted by

Anup chaudhary [KAN075BCT009]

Chris Gurung [KAN075BCT023]

Himalaya Pal [KAN075BCT029]

Kundan Giri [KAN075BCT030]

in partial fulfillment for the degree of Bachelor in Computer Engineering.

.....
Supervisor
none
Supervisor’s Designation
Second Line of Designation (if required)

.....
External Examiner
External’s Name
External’s Designation
Second Line of Designation (if required)

.....
Er. Rabindra Khati
Head of Department
Department of Computer and Electronics Engineering

Date: June 12, 2022

ABSTRACT

Data security is a crucial concern that ought to be managed to help protect vital data. Cryptography is one of the conventional approaches for securing data and is generally considered a fundamental data security component that provides privacy, integrity, confidentiality, and authentication.

In the current world where communication has been made easy such that you could talk to a person on the other side of the world with a press of the button. With the increase in availability of internet service you can send texts, photos ,files through the internet in a matter of seconds and for far less cheaper. This is achieved through different chat applications. With the increased usage of such chat applications the contents of such messages contains more that just simple messages to friends and families but also very important information and files which on the wrong hands could cause a huge catastrophe. As such End-to-End security is needed to safely exchange private information with each other without worrying about data. With this project we aim to provide an End-to-End encrypted chat apps with file compression feature. List of requirements to make such application are provided in this paper.

This project approach End-to-end encryption method to provide secure communication that prevents third parties from accessing data while it's transferred from one end system or device to another. The end-to-end encryption method is implemented using the Asymmetric key encryption algorithm (RSA). We compressed the message using huffman Encoding algorithm. Then used RSA to encrypt that encoded data. The login system is also secured because we used AES library to secure the password.

In this way we can maintain the data more securely. Since we used huffman to compress the data and RSA algorithm for securing the data.

Keywords – RSA, Huffman

ACKNOWLEDGMENT

We would like to express our gratitude and appreciation to all those who gave us the possibility to complete this report. Special thanks is due to our project supervisor Mr. Bishal Thapa sir whose help, stimulating suggestions and encouragement helped us in all time of fabrication process and in writing this report. We also sincerely thanks for the time spent proofreading and correcting our many mistakes

Many thanks go to the all lecturer and supervisors who have given their full effort in guiding the team in achieving the goal as well as their encouragement to maintain our progress in track. Our profound thanks go to all classmates, especially to my friends for spending their time in helping and giving support whenever we need it in fabricating my project

Anup chaudhary	[KAN075BCT009]
Chris Gurung	[KAN075BCT023]
Himalaya Pal	[KAN075BCT029]
Kundan Giri	[KAN075BCT030]

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

With the rapid development of mobile devices, computers and accessibility to the internet there is growing users of different chat applications. The attracting or more so important features of any social media has become the chat feature. Such chat features provide real time messaging, file sharing which may include different photos, videos, documents, etc. Chatting has been such an important aspect of the modern world that it is expected the no. of active users currently is in billions with this number expected to increase more in the coming years. Among some of the popular chat applications Whatsapp's monthly active users(in millions) are 2000, messenger 988, snapchat 557 to name a few.[1]

The importance of such apps was never more clearly presented as when Facebook experienced outage in 2021. In what was just a 6 hour long outage the competing apps like telegram gained a record 70 million new users. [2] As the dependence on chat systems grow day by day the vulnerability and assaults also increases. As such there is increasing need to implement a secure communication

Digital communication witnesses a noticeable and continuous development in many applications in the Internet. Hence, secure communication sessions must be provided. The security of data transmitted across a global network has turned into a key factor on the network performance measures. So, the confidentiality and the integrity of data are needed to prevent eavesdroppers from accessing and using transmitted data.

One way of to secure the communication is using End-To-End encryption which use the Cryptographic algorithm. Cryptographic algorithms are divided into symmetric and asymmetric keys. The symmetric algorithms require a single key only for the encryption and decryption of data. Asymmetric algorithms on the other hand require both public and private keys for the encryption and decryption of data. The scrambling of the data is done using the public key while the private key is made known only to the

receiver which is meant for the decryption of the data. A maiden asymmetric algorithm was proposed by Diffie-Hellman [2][3] which ensures secured communication as well as data security. A counter algorithm termed RSA which has lower time complexity based on prime number factorization was proposed in 1977 and is the patent of Ron Rivest, Adi Shamir, and Len Adleman (RSA), which was published in 1978 at the Massachusetts Institute of Technology [4]. In this algorithm, two prime numbers are used to produce the public and the private key. When the keys are created, the prime numbers are no more considered and are or can be discarded.

1.2 Problem statement

The purpose of this project is to provide the correct data with security to the users. For some of the users the data might be lost during the transmission process in the network and for some, the data might be changed by the unauthorized person in the network and there are some other security problems in the network. Our application will give you more Security to the data present in the network and there will be able to reduce the loss of data in the network which will be transmitted from the sender to the receiver using the latest technologies. Only the Authorized persons i.e., who are using our application will be there in the Network.

The proposed algorithm is to compress the message at one end and ask the public key of another person and use that public key to encrypt the message and the person only decrypt the message using his private key.Using Huffman encoding the message compressed is loss less.

1.3 Objectives

The application aims to provide data security in communication system. The application focuses on simplicity of design, having user-friendly interface and to be easily understood.

The main objectives of the application can be enumerated as follows:

- To provide end-to-end data security.
- To provide platform for sending messages from one person to another.

1.4 Applications

The application is an online web application. This system can be used to provide end to end encryption of data and also used to provide secure connection between user between users with smooth and clean UI.

1.5 Project features

The application, is targeted towards the general population, so the core features of this applications can be listed into two categories depending on whether user wants to view the predicted values or just access the inventory management as follows:

1.5.1 Value predictions

- Shows entire list of produces and predicted prices,
- Allow user to select the predicted price according to day, month or year.

1.5.2 Inventory management

- Allow user to add, delete and update inventory,
- Allow user to create a profile and save their inventory,

1.6 Feasibility Analysis

1.6.1 Economic Feasibility

Based on our economic analysis for development and operational cost, the system is being developed and operated economically. For development, the required devices are readily available, so it is feasible. Also, it is economically feasible to the consumers as it costs no charge to use the platform.

1.6.2 Schedule Feasibility

Based on the objectives and the time left for the development. The schedule is found to be feasible.

1.6.3 Technical Feasibility

Technically, the system is feasible enough and easy-to-use for both technical and non-technical groups of people. It provides a user-friendly environment along with features

using the latest technologies. The system provides a layout most of the applications people are used to anyway so it will be easy to use.

1.6.4 Operational Feasibility

For the operation of the system, the person does not need to excel in using a computer. Since, the event may not always be related to the technical fields, someone with minimum knowledge about computer and technology can also get benefit from the system. Similarly, one can get access to the system as a web-based application. There is no requirement of huge and expensive hardware. The system comprises only of farmer's end.

1.7 System Requirements

1.7.1 Software Requirement

Application is targeted towards a general market, so it is aimed to be fully optimized enough for any low-range to high-range systems, so listed below are the software requirements for the development and operation of this system:

- Operating System: Windows 8 or above
- Browsers: Google Chrome, Firefox, etc
- PostgreSQL Server
- Python

1.7.2 Hardware Requirement

Hardware configuration and requirements for the operation of this application are as follows:

- Intel Core 2 Duo Processor (Recommended i-series processors or more) with minimum of 2GB RAM for application operation
- Server with optimum node speed

CHAPTER 2
LITERATURE REVIEW

Applications of Machine Learning Techniques in Agricultural Crop Production describes about various Machine Learning applications that would prove to very useful in the agriculture sector. For these applications, a large amount of data available from many resources can be analyzed to find the hidden knowledge. This research field is growing day by day and will prove to be a great tool for the development of the agriculture sector in the future. The combination of Agriculture and Computer Science will provide a great scope of development in the agriculture sector. The paper, Smart Farming: A Techno Agriculture Advancement Powered by Machine Learning's main aim is to make the agriculture sector aware about the modern technologies. Accurate predictions should be made with the help of machine learning instead of manual predictions so as to improve the commercial value of the crops. The main problem identified in the paper is that Nepal is the only country which lacks in technological advancements in the agriculture sector, due to which manual predictions are done for everything. These results in crops being sold at a less price and sometimes even the crops are ruined due to wrong predictions about the weather. The solution to this is Machine Learning. ML is the technique to provide knowledge to the machine so that it can think on its own provided with the correct data. This will help to raise the standards of agriculture in Nepal. Machine Learning in Agriculture. This paper, Machine Learning in Agriculture: A Review is based takes a practical approach and implements many learning models and algorithms in the field of Agriculture. The paper, Crop Price Prediction System using Machine Learning Algorithms' main objective is to estimate the crop price by analyzing the existing data using certain data analytics techniques. This paper shows a more of a practical approach towards the topic. Data from various sources have been collected and a system is created for the crop price prediction. The whole system has been implemented using the python programming language. The data is collected from reliable sources and stored in a storage where it is then used accessed, transfer and analyzed by an organization. The data is then processed and transform the raw data into a more efficient format. Machine learning techniques like Decision tree regression is used here to determine important information and to increase the accuracy percentage of the

price prediction. The final results are shown through visual elements like bar graphs. This paper is very useful to understand the importance of machine learning techniques and how modern technologies can prove to be very useful in the development of agriculture sector. This paper, Crop Price Prediction using Decision Tree Regression's main objective is to predict the price of the crop and estimate the profit for the crops given in the system before sowing. The databases provide enough data for predicting the appropriate Market Selling Price for the crops and their demand in the market.This paper shows that the Decision tree regression is a very effective technique in the prediction. As there are many algorithms in machine learning in agriculture, there are various ways to make predictions for the crops which is also one of the main things this paper has shown, that there are different algorithms for different crops and not a single one for all the crops. Thus, this paper shows how machine learning can prove to be beneficial for the farmers to take the right decisions in choosing the crops by analyzed results.[1]

CHAPTER 3
METHODOLOGY

3.1 Required Algorithm:

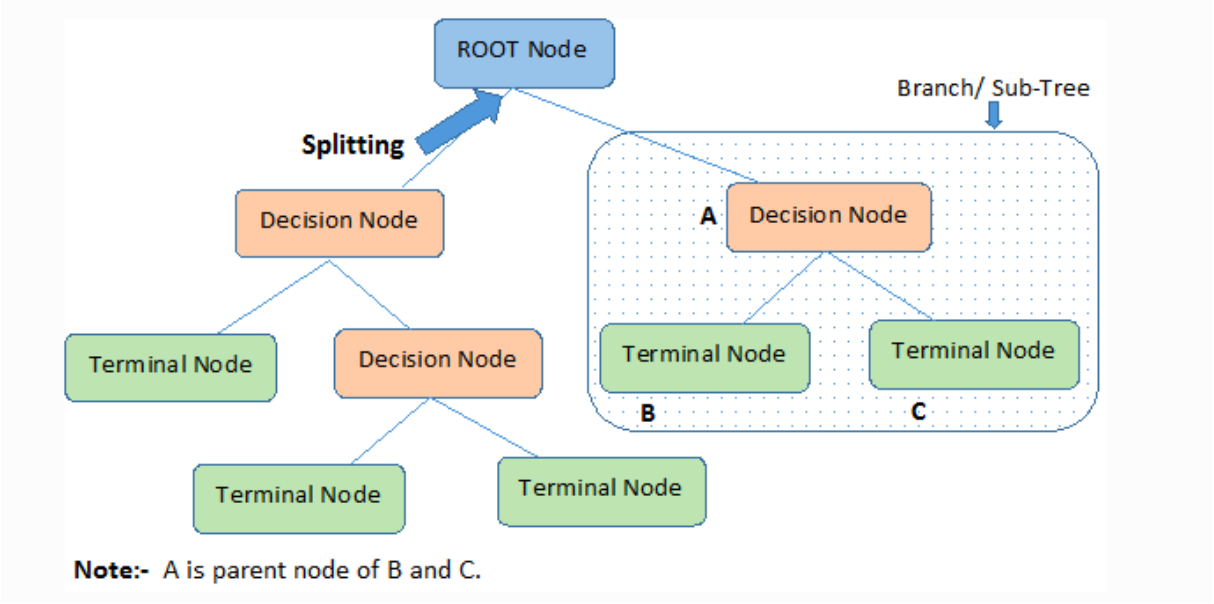
3.1.1 Overview of Decision tree Algorithm

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

1. **Important Terminology related to Decision Trees**
    - **Root Node:** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
    - **Splitting:** It is a process of dividing a node into two or more sub-nodes.
    - **Decision Node:** When a sub-node splits into further sub-nodes, then it is called the decision node.
    - **Leaf / Terminal Node:** Nodes do not split is called Leaf or Terminal node.
    - **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
    - **Branch / Sub-Tree:** A subsection of the entire tree is called branch or sub-tree.
    - **Parent and Child Node:** A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.



Decision trees classify the examples by sorting them down the tree from the root to some leaf/terminal node, with the leaf/terminal node providing the classification of the example.

Each node in the tree acts as a test case for some attribute, and each edge descending from the node corresponds to the possible answers to the test case. This process is recursive in nature and is repeated for every subtree rooted at the new node.

2. **Assumptions while creating Decision Tree**

   Below are some of the assumptions we make while using Decision tree:

   - In the beginning, the whole training set is considered as the root.
   - Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
   - Records are distributed recursively on the basis of attribute values.
   - Order to placing attributes as root or internal node of the tree is done by using some statistical approach.

   Decision Trees follow Sum of Product (SOP) representation. The Sum of product (SOP) is also known as Disjunctive Normal Form. For a class, every branch from the root of the tree to a leaf node having the same class is conjunction (product) of values, different branches ending in that class form a disjunction (sum).

   The primary challenge in the decision tree implementation is to identify which

attributes do we need to consider as the root node and each level. Handling this is to know as the attributes selection. We have different attributes selection measures to identify the attribute which can be considered as the root note at each level.

3. **Reduction in Variance**

   Reduction in variance is an algorithm used for continuous target variables (regression problems). This algorithm uses the standard formula of variance to choose the best split. The split with lower variance is selected as the criteria to split the population:

   Steps to calculate Varience and varience reduction:

   (a) Calculate variance for each node.

   (b) Calculate variance for each split as the weighted average of each node variance.

   (c) Differentiate varience of child from varience of parent.

   Used Formulae:

   Average = $\frac{\sum x}{n}$

   Standard Deviation(Sd)=$\sqrt{\frac{\sum x-\bar{x}}{n}}$

   Varience(var)=$\frac{1}{N}\sum\left(x_i-\bar{x}\right)^2$

   Varience Reduction= $var(parent)-\sum w_i var(child_i)$

3.1.2 Data Preprocessing

Step 1:Importing the required libraries

These two are essential libraries which we will import every time. Numpy is a library which contains mathematical functions. Pandas is the library used to import and manage the data sets.

Step 2:Importing the data set

Data sets are generally available in .csv format. A CSV file stores tabular data in plain text. Each line of the file is a data record. We use the read.csv method of the pandas

library to read a local. CSV file as a data frame. Then we make separate Matrix and Vector of independent and dependent variables from the dataframe.

Step 3: Handling the Missing Data

The data we get is rarely homogeneous. Data can be missing due to various reasons and needs to be handled so that it does not reduce the performance of our machine learning model. We can replace the missing data by the Mean or Median of the entire column. We use Imputer class of sklearn.preprocessing for this task.

Step 4: Encoding Categorical Data

Categorical data are variables that contain label. values rather than numeric values. The number of possible values is often limited to a fixed set. Example values such as "Yes" and "No" cannot be used in mathematical equations of the model so we need to encode these variables into numbers. To achieve this we import Label Encoder class from sklearn.preprocessing library.

Step 5: Splitting the dataset into test set and training set

We make two partitions of dataset one for training the model called training set and other for testing the performance of the trained model. called test set. The split is generally $80/20$. We import train test split method of sklearn.crossvalidation library.

Step 6: Feature Scaling

Most of the machine learning algorithms use the 71 Euclidean distance between two data points in their computations, features highly varying in magnitudes, units and range pose problems. High magnitudes features will weigh more in the distance calculations than features with low magnitudes. Done by Feature standardization or Z-score normalization. Standard Scalar of sklearn.preprocessing is imported.

3.1.3 Steps for Decision Tree Regression Model:

Step 1:Importing the libraries

The first step will always consist of importing the libraries that are needed to develop the ML model. The NumPy, matplotlib and the Pandas libraries are imported.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
```

Step 2:Initialize and print the Dataset

```
data = pd.read_csv("Potato_cleaned_data.csv")
data.head(5)
```

	Unnamed: 0	index	Commodity	Date	Unit	Minimum	Maximum	Price	Day	Month	Year
0	0	3	Potato White	2013-06-16	Kg	15.0	16.0	15.5	16	6	2013
1	1	77	Potato White	2013-06-17	Kg	15.0	16.0	15.5	17	6	2013
2	2	150	Potato White	2013-06-18	Kg	15.0	16.0	15.5	18	6	2013
3	3	223	Potato White	2013-06-19	Kg	15.0	16.0	15.5	19	6	2013
4	4	296	Potato White	2013-06-20	Kg	15.0	16.0	15.5	20	6	2013

Step 3: Select all the independent features from the dataset to “X” and dependent features to ”y”

```
X = data[["Day","Month","Year"]].values
Y = data["Price"].values.reshape(-1,1)

x_train, x_test, y_train, y_test = train_test_split(X,Y,train_size=0.8,shuffle=True, random_state=1)
x_train.shape
x_test.shape
```

Step 5: Fit decision tree regressor to the dataset

```
X = data[["Day","Month","Year"]].values
Y = data["Price"].values.reshape(-1,1)

x_train, x_test, y_train, y_test = train_test_split(X,Y,train_size=0.8,shuffle=True, random_state=1)
x_train.shape
x_test.shape
```

```
(477, 3)
```

Step 6: Predicting a new value

```
y_predict = regressor.predict(x_test)
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import MinMaxScaler,StandardScaler
np.sqrt(mean_squared_error(y_test, y_predict))
```

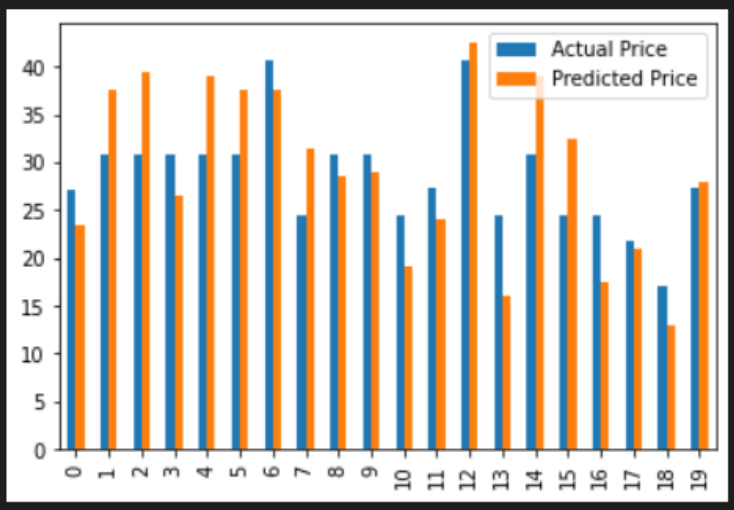
```
5.870662989923151
```

Step 7: Visualising the result

```
import matplotlib.pyplot as plt
graph = dfr.head(20)
graph.plot(kind='bar')
```

✓ 0.9s

<AxesSubplot:>



3.2 Software development model

3.2.1 Incremental Model

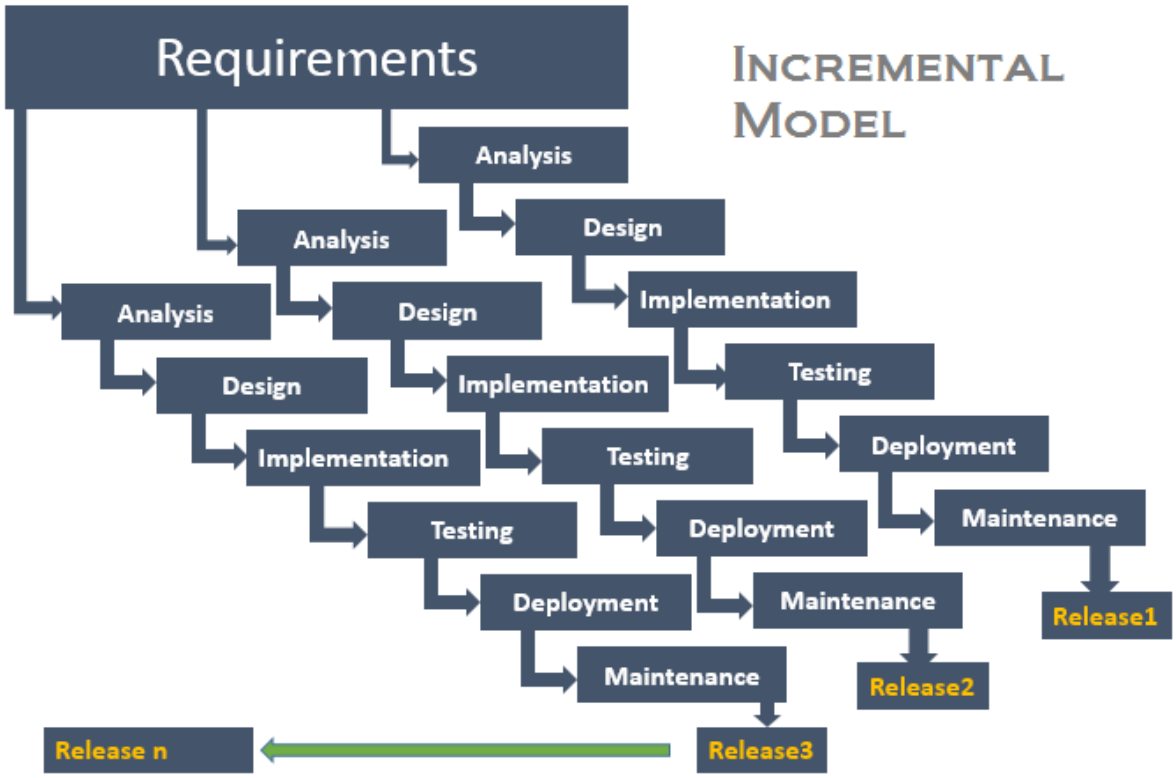


Figure 3.1: Incremental Model Block Diagram

First Increment:

First, the website was analyzed to know how it should look like. Then, the designing of templates was done. After observing the design of the templates, we started coding using hypertext languages Html, CSS, JavaScript, and Bootstrap. And finally, after testing the frontend was created.

Second Increment:

After analyzing the scenario of the project, the algorithms to be implemented was analyzed. Using the algorithms, we started designing the algorithms that is suitable for the project. Initiating the coding we completed the algorithm implementation. The testing

of algorithm was done using the dataset. Finally, the algorithm was implemented.

Third Increment:

After the algorithm implementation backend designing and coding was started. Using Django and Python the backend part was completed and tested. Finally, backend was ready.

Fourth Increment:

Finally, after all the designing was completed, coding for the project was done. Later, testing of the project was done. At last, the final website was designed.

3.3 Block Diagram:

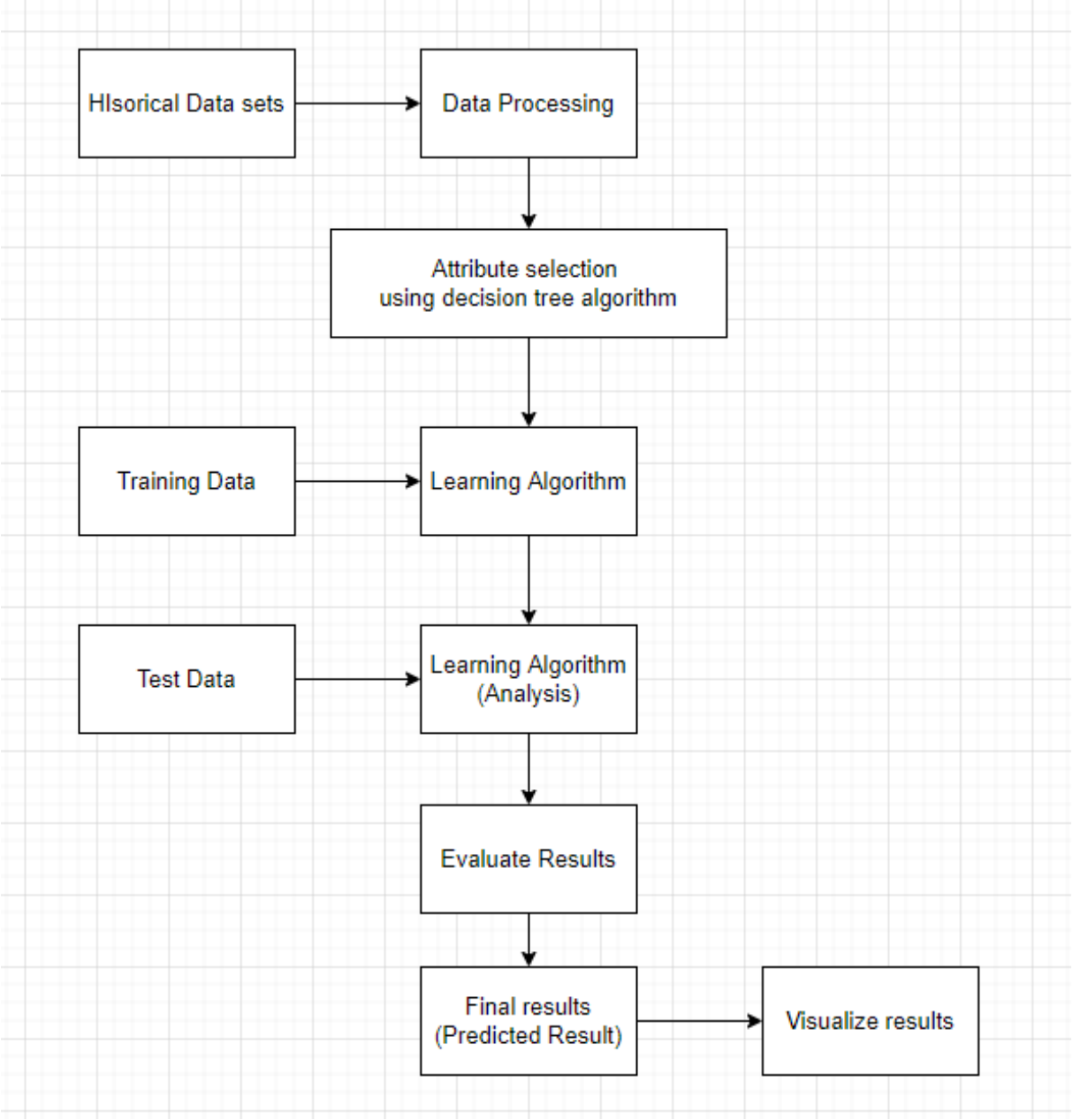


Figure 3.2: Block Diagram

3.4 Use Case Diagram:

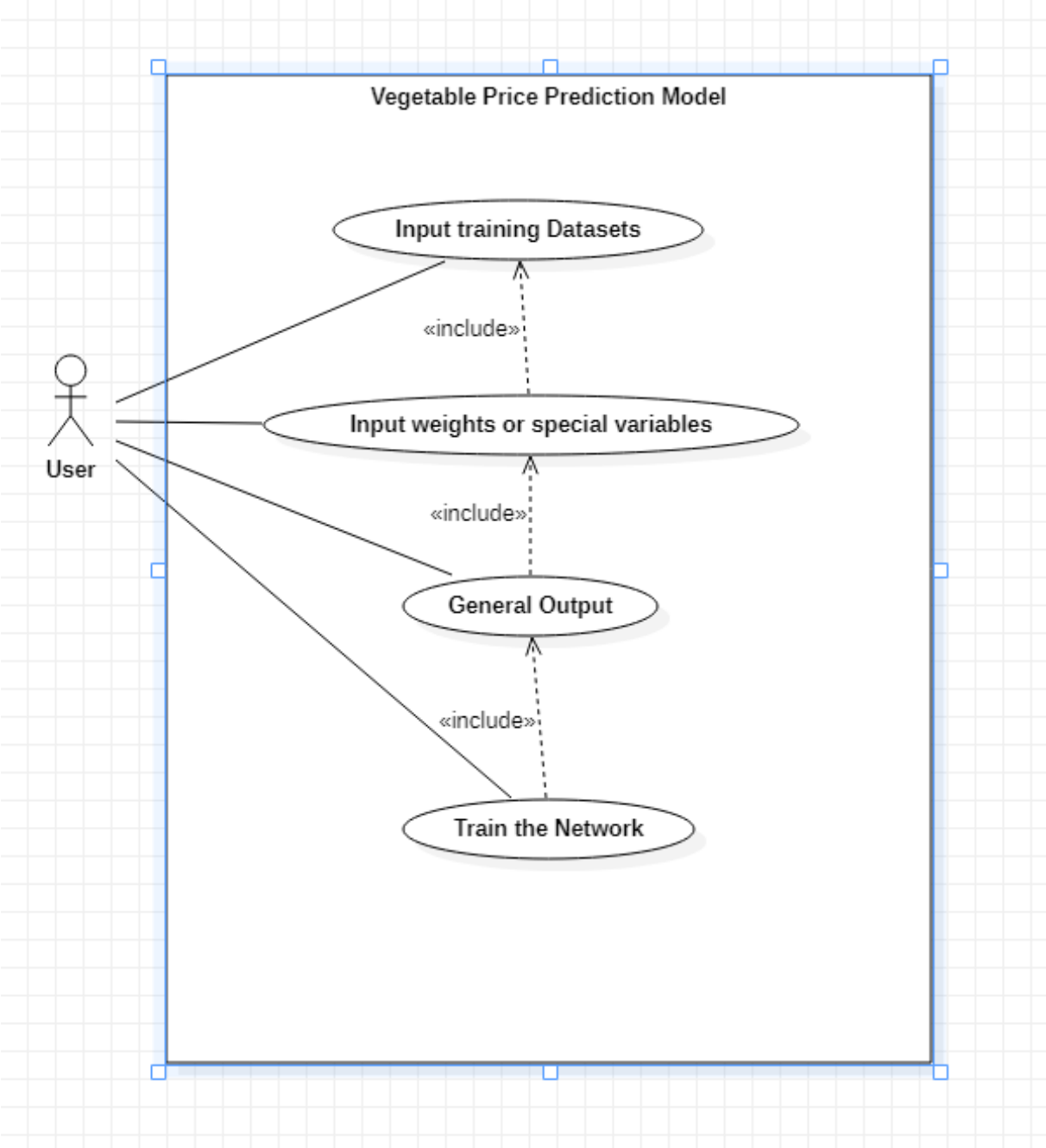


Figure 3.3: Use Case Diagram

CHAPTER 4
RESULT AND DISCUSSION

4.1 Output:

We succesfully created a vegetable price predictiction model which can predict price with good accuracy. All the the data to train model were taken from Kalimati market price dataset to complete our project. We use data from year 2013 t0 2021 and use hundreds of dats's for better accuracy of our model.

We then trained all the datasets to make a model. We tried to create efficient and accurate model to identify the vegetable and predict it's price of the given date. After training and testing of model, a user friendly web UI is made for the user to make task easier.

Machine learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a data-set based on the training data.

In Machine Learning, error is used to see how accurately our model can predict on data it uses to learn; as well as new, unseen data.

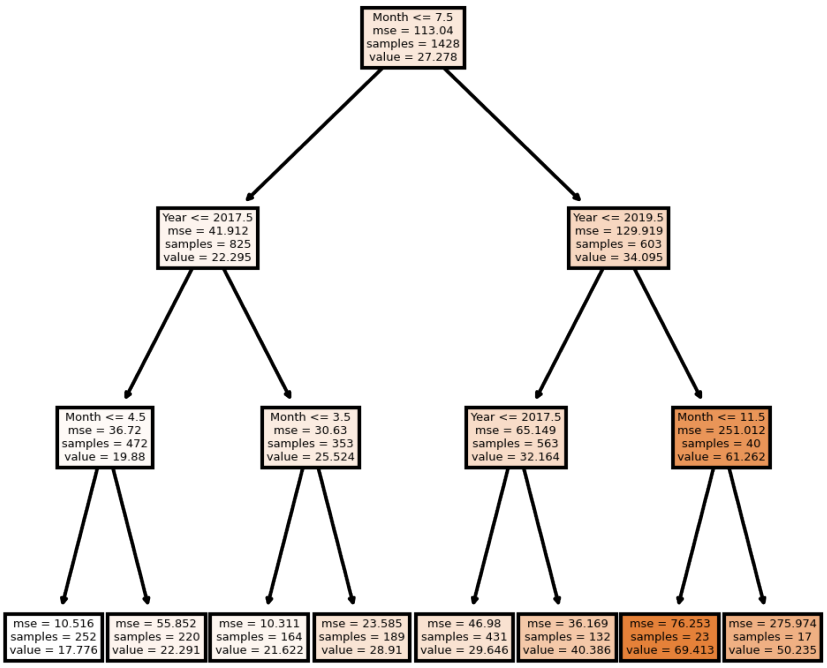


Figure 4.1: Vegetable Price Prediction Decision Tree

The above figure shows the regression tree working for our model. This figure includes year, month, regreesion value etc which shows how regression tree method is working and providing us assist for creating vegetable price prediction model.

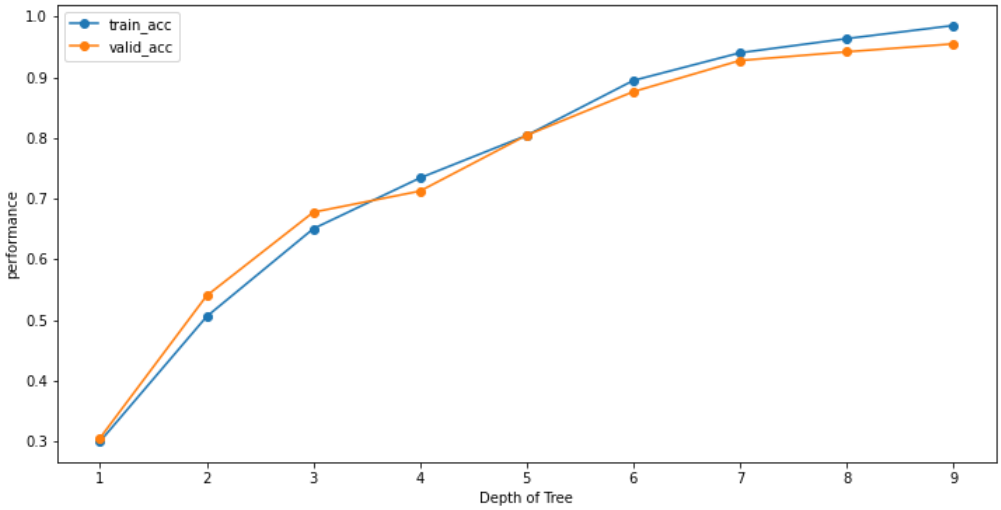


Figure 4.2: Vegetable Price Prediction Model Accuracy

The above figure shows the accuracy of our model. It has accuracy of about 75%.

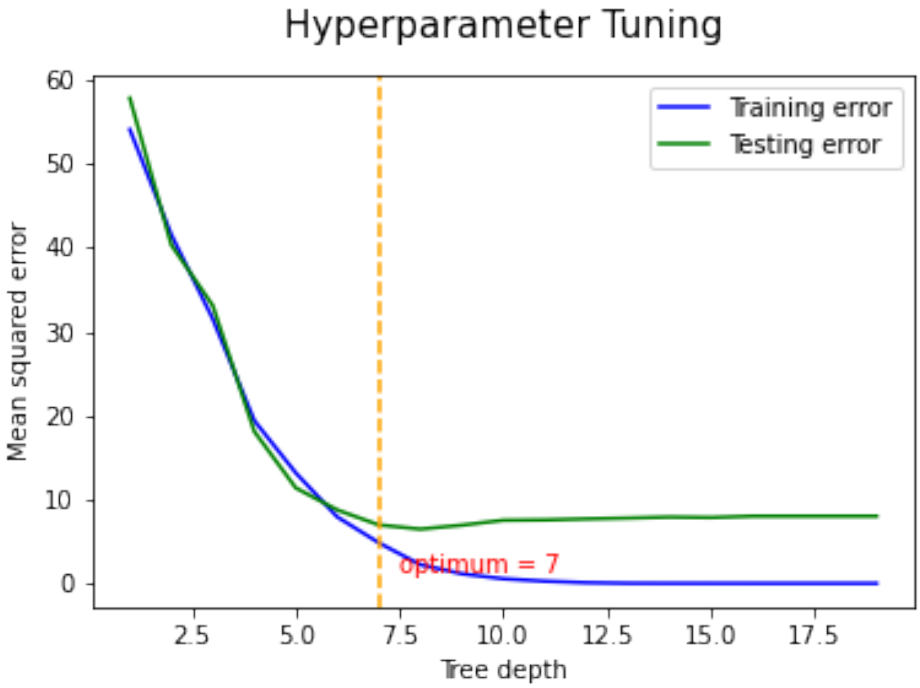


Figure 4.3: Vegetable Price Prediction Mode error

The above figure shows the error of our model. We get mean square error of approximate 34.46, absolute error of approx 4.34 and mean square error of about 5.87.

4.2 UI output

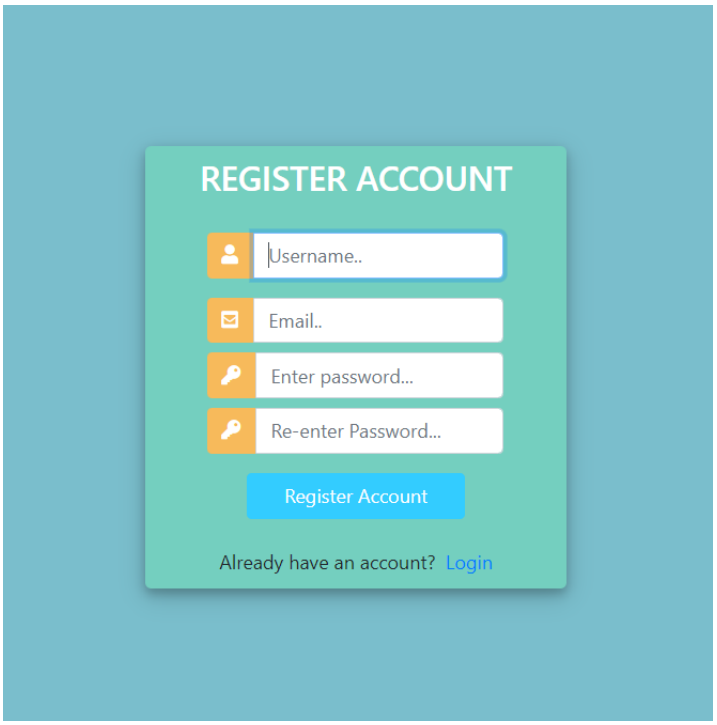


Figure 4.4: Sign up interface

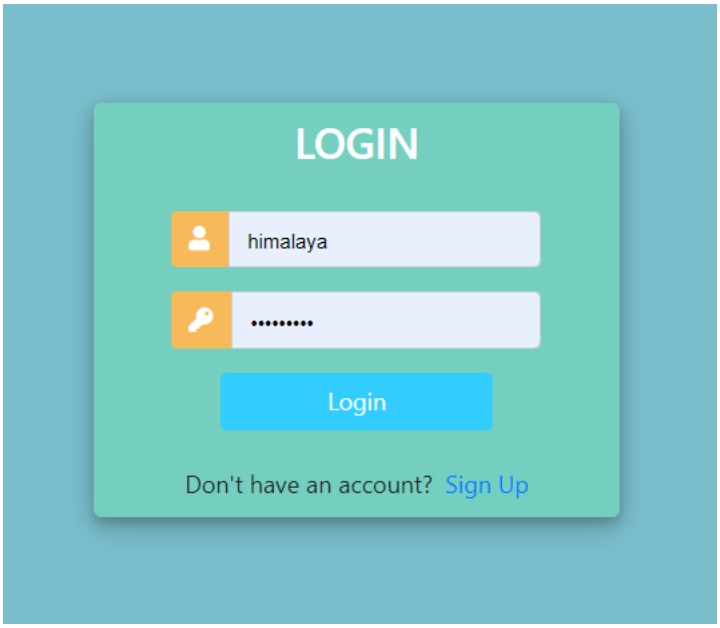


Figure 4.5: Sign in interface

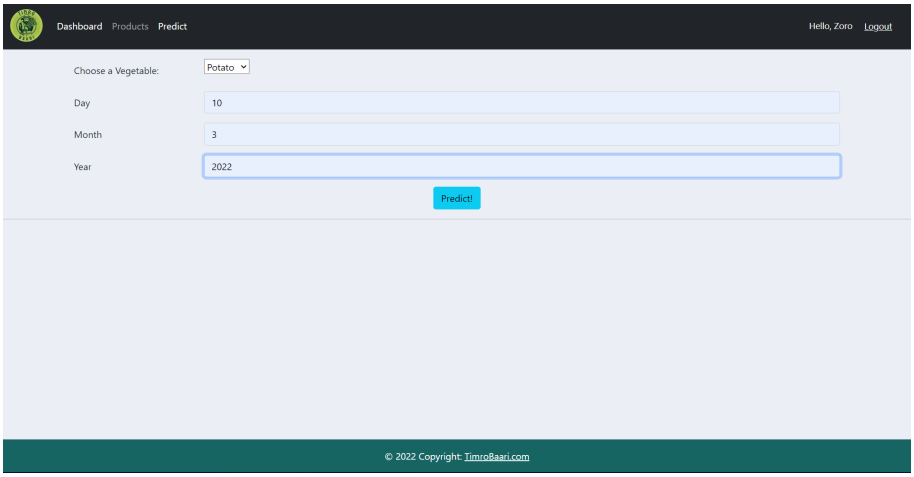


Figure 4.6: Output of UI

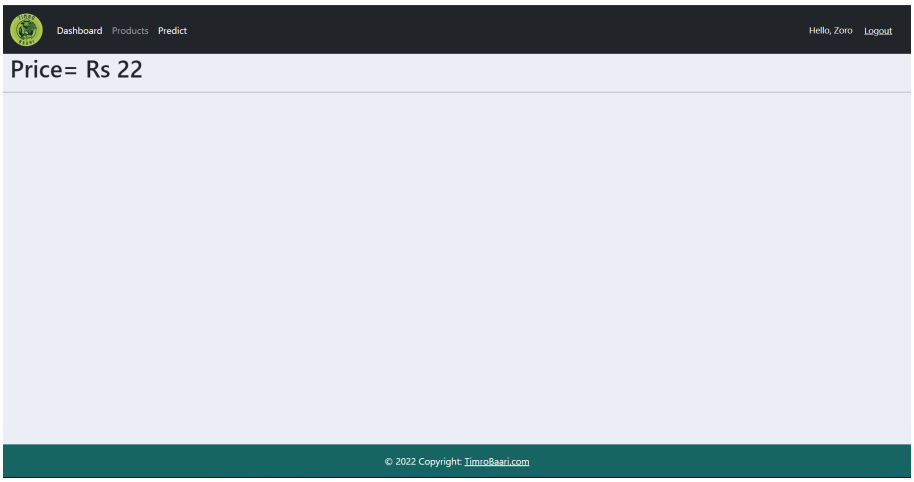


Figure 4.7: Output of UI

The above figures shows the User interface of our website where we can sign up , log in and predict price of vegetables using day , month and year.

CHAPTER 5
CONCLUSION

Timro Baari (A vegetable Price Prediction Model) uses machine learning method using regression tree. It can detect the price of vegetables of given date (year/month/date) of vegtable which we choose to predict. This can help different aspects of agriculture and farmers can predict price through which they can mass produce vegetables in which they will get more profit. As well as costumer can predict price of vegetables and buy more vegetables which they need before increment of price. It will help in production of vegetables of farmers in which they will not have loss. This proposed model has accuracy of about 75

5.0.1 Possible Future research

In future we can use other various types of machine learning method for better accuracy of prediction. We can also add tremendous amount of datasets and more type of vegetables choice. We can also make our Web UI more user friendly and clean.

REFERENCES

[1] M. Rakhra, P. Soniya, D. Tanwar, P. Singh, D. Bordoloi, P. Agarwal, S. Takkar, K. Jairath, and N. Verma, "Crop price prediction using random forest and decision tree regression:-a review," *Materials Today: Proceedings*, 2021.