CROP PRICE PREDICTION

Project Report submitted to

Guru Jambheshwar University of Science and Technology, Hisar

For the partial award of the degree

Of

Master of Computer Applications

By

Abhishek (200010120001)
Abhishek (200010120002)
MCA Final Year
GJUS&T, Hisar

Dr. Amandeep Noliya
Assistant Professor
Department of CSE
GJUS&T, Hisar

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GURU JAMBHESHWAR UNIVERSITY OF SCIENCE AND TECHNOLOGY, HISAR

July-2022

DECLARATION

We, Abhishek, 200010120001 and Abhishek, 200010120002 certify that the work contained in this project report is original and has been carried by us under the guidance of Dr. Amandeep Noliya. This work has not been submitted to any other institute for the award of any degree or diploma and we have followed the ethical practices and other guidelines provided by the Department of Computer Science and Engineeringin preparing the report. Whenever we have used materials (data, theoretical analysis, figures, and text) from other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references. Further, we have taken permission from the copyright owners of the sources, whenever necessary.

Signature	Signature
Abhishek	Abhishek
200010120001	200010120002
Department of CSE	Department of CSE
GJUS&T, Hisar	GJUS&T, Hisar

Signature

Supervisor:-**Dr. Amandeep Noliya**

Designation: - **Assistant Professor**

Department of Computer Science and Engineering

Guru Jambheshwar University of Science and Technology, Hisar

CERTIFICATE

This is certified that Abhishek (200010120001), Abhishek (200010120002) has worked under my supervision to prepare his project on **"Crop Price Prediction"**. They have worked on their project through the semester from March 2022 to JULY 2022.

I wish him success in life.

Dr. Amandeep Noliya
Assistant Professor
Department of CSE
GJUS&T, Hisar

PLAGIARISM CERTIFICATE

This is certify that Abhishek (200010120001) , Abhishek (200010120002) are students of MCA(CSE) , Department of computer Science & Engineering , Guru Jambheshwar University of Science & Technology, Hissar has completed the project entitle **“Crop Price Prediction”**.

This complete project report has been checked by Turnitin Software and the similarity index is **<u>8%</u>** i.e. the accepted norms of university. The project report may be considered for the award of the degree.

Signature:
Abhishek (200010120001)
MCA Final Year
Department of CSE, GJUS&T, Hisar

Signature:
Abhishek (200010120002)
MCA Final Year
Department of CSE, GJUS&T, Hisar

ACNOWLEDGEMENT

An assignment puts to the litmus test an individual’s knowledge, credibility, and experience and thus sole efforts of an individual are not sufficient to accomplish the desire. Successful Completion of the project involves interest and efforts of many people. So, this becomes Obligatory on my part to record my thanks to all of them.

Therefore, in this way would like to thank **Dr. Amandeep Noliya**. The project supervisor, who had allowed us to work and listened to our progress report timely, He had discussed about the project and suggested us to work in enhanced manner.

At last, but not least, we express our heartiest gratitude to all our friends and Supervisor and all in all almighty for providing us a favourable environment and support.

Signature:
Abhishek (200010120001)
MCA Final Year
Department of CSE, GJUS&T, Hisar

Signature:
Abhishek (200010120002)
MCA Final Year
Department of CSE, GJUS&T, Hisar

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

ABSTRACT

The foundation of the Indian economy is agriculture. A little more than 60% of people are depending on agriculture and 95% of its doesn't know that which crop is beneficial to grow according to price. Farmers aren't able to get relevant price for their crops. As we know the price of a crop is depend on supply and demand but due to lack of education or resource we can't predict or estimate the value. Many researches are done in this field but most of the researches are based on yield of land according to yield and temperature they predict the value and quantity. The main drawback of the previous researches is that they work on the previous data but the objective of this paper is that the farmer get an estimated value before they grow we try to work on real time data. "According to 'Adam Smith' The law of supply and demand is a theory that explains the interaction between the sellers of a resource and the buyers for that resource. The theory defines the relationship between the price of a given good or product and the willingness of people to either buy or sell it. Generally, as price increases, people are willing to supply more and demand less and vice versa when the price falls". With the use of this theory and machine learning we design a model which gives us the predict value. This prediction helps the farmer and the government also to make efficient decisions.

In this project, we implement a supervised machine learning algorithm namely Decision tree, Linear Regression, Multi – Linear Regression, KNN Regression, Gradient Boosting,

Keywords: Crop-yield, Supervised Machine learning in Agriculture, Decision tree, Forecasting, Crop price prediction based on demand and supply.

Chapter 1

Introduction

1.1 Overview

With more than 1.6 million square kilometers of the second-highest land area under agriculture [20] India is an agricultural nation. The foundation of the Indian economy is agriculture. Around 60% of the population is dependent on agriculture and 95% of its doesn't know that which crop is beneficial to grow according to price. As per the 2018 report India 60% are employed in agriculture and 17 – 18 % are in an Indian growth economy.

Every farmer in the globe deals with the economic issue of crop price fluctuation. Farmers aren't able to get relevant price for their crops. Farmers are unable to predict how the market's supply and demand will balance out in the future. As we know the price of a crop is depend on supply and demand but due to lack of education or resource we can't predict or estimate the value. There is a chance of selling an agricultural product of inferior quality as a result.

In this regard, the majority of choices are made at the start of the planting season, however not all of them are changeable. As a result, price predicting is an important step, and crop price forecasting research is very important.

The price of many agricultural products varies widely as a result of the knowledge asymmetry between production and marketing. Numerous elements, including climate, pricing, policy, and others, have an impact on crop prices, can act as a social cue for the supply and demand of goods. A variety of factors, including the season, geography, economic status, and social factors, frequently affect the prices of agricultural products on the market. Direct indicators of crop supply and demand can be found there. According to 'Adam Smith' "The law of supply and demand is a theory that explains the interaction between the sellers of a resource and the buyers for that resource. The theory defines the relationship between the price of a given good or product and the willingness of people to either buy or sell it. Generally, as price increases, people are willing to supply more and demand less and vice versa when the price falls"[1].

There is a special connection between its crop price to the demand for that specific crop and production. If the demand is high then there is more chance that farmers get high prices. But in case of low demand and high production, there are more probability farmers not get the base price even not cover the expenses.

So, there is a gap between the farmer and government, if there is an opportunity for the farmer to gain the right or high price of his price, then we believe they grab it if there is any platform that suggests the farmer which crop will he have to sow in the field and which not. In order to make informed judgments and ensure that the social economy is operating favourably, agricultural authorities must have a precise forecast of the price of agricultural commodities.

Government has the estimated demand stats if the farmer update which crop sow or want to sow, then our portal provide the info to the customer that which crops his better to sow and not.

Many researches are done in this field but most of the researches are based on yield of land according to yield and temperature they predict the value and quantity. The main drawback of the previous researches is that they work on the previous data but the objective of this paper is that the farmer get an estimated value before they grow we try to work on real time data.

The nonlinear planning in this study has been tested using historical pricing data. Planning for the future, however, requires a long-term or short-term price estimate. Short-term and long-term predictions serve various functions. Even though the produce is stable because to greenhouses, farmers' revenue is still not steady. This is so that their production can't respond as fast to market demands as industrial products can.

We will go to train our model which gives the demand on basis of last year's production of that crops or variety price. Our model gives the demand, which will also be a parameter to calculate the related price of that specific crop.

With the use of this theory and machine learning we design a model which gives us the predict value. This prediction helps the farmer and the government also to make efficient decisions.

This is portal suggest the selection of crop for sowing on basis of the current time scenario of crops and demand.

1.2 Machine Learning

A subset of artificial intelligence is called machine learning. The main objective of machine learning is to make machines as smarter as humans. Machine learning makes a decision and executes the task without the interaction of humans, the machine takes decisions on the previous data and learning. The term machine learning was first introduced by **Arthur Samuel** in **1959**.

Because drawing inferences from a sample is its primary function, machine learning builds mathematical models using the theory of statistics. The two stages of machine learning methods are as follows:

1. Training: From a set of training data, a model is learned.
2. Use: Decisions concerning some fresh test data are made using the model.

Tom Mitchell provides definition “If a computer program's performance at tasks T, as evaluated by performance measure P, increases with experience E, then it is said to have learned from experience E with respect to that class of tasks T and performance measure”.

With the aid of machine learning, a machine may predict outcomes without being explicitly programmed and automatically learn from data. Computers now have the ability to automatically learn from prior data thanks to an emerging technology. In order to build mathematical models and generate predictions based on previously collected data or information, machine learning employs a range of methodologies. Currently, It is employed for a variety of purposes, including speech recognition., recommender systems, email filtering, Facebook auto-tagging, and image identification.

With the aid of previous sample data, sometimes known as "training data," machine learning algorithms build a mathematical model that helps them make predictions or judgments without being explicitly programmed. Computer science and statistics[18] are used with machine learning to create prediction models. Algorithms that learn from past data are created by machine learning or used in it. The performance will be higher the more information we supply. If additional data can be gathered to help a machine perform better, it can learn.

1.2.1 Features of Machine Learning:

1. Data is used by machine learning to find different patterns in a dataset.
2. Resolving complex issues that are challenging for people to solve.
3. Making decisions in a variety of fields, including finance.
4. Analyzing data for hidden patterns and important information.

1.2.2 Applications of machine learning

1. Machine learning is utilized in the retail industry to analyze consumer behavior.
2. Banks look at their historical data to develop models for the stock market, fraud detection, and loan applications in the realm of finance.
3. Production uses learning models for control, optimization, and troubleshooting.
4. In medicine, learning programs are used to make diagnosis.
5. Call patterns are examined in telecommunications to optimize the network and increase service quality
6. Only computers have the processing power to analyze enormous amounts of data in physics, biology, and astronomy with sufficient speed. It is impossible to manually search the vast and constantly developing World Wide Web for relevant information.
7. A system is trained to learn and adapt to changes, relieving the system designer of the burden of anticipating every scenario and coming up with a solution.
8. It is utilized to solve a variety of issues in robotics, speech recognition, and vision.
9. Computer-controlled vehicles are designed using machine learning techniques to steer accurately on a variety of roads.
10. Chess, backgammon, and Go playing programs have been created using machine learning techniques.

1.3 Components of Learning

The fundamentals of the learning process the four parts of learning, whether done by a human or a machine, are data storage, abstraction, generalization, and evaluation. The numerous elements and phases involved in the learning process are illustrated in figure 1.1.

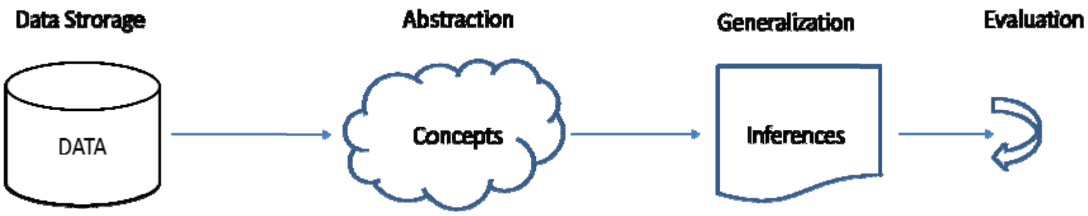


figure1.1: Components of learning process

1. ***Data storage:***A crucial element of the learning process is the availability of facilities for storing and retrieving enormous volumes of data. Data storage serves as the basis for sophisticated thinking in both humans and computers.

   Computers employ hard disc drives, random access memory, flash memory, and similar devices to store data and use cables and other equipment to retrieve data. In a person, Electrochemical impulses are used to extract data from the brain's storage units.

2. ***Abstraction:***The second component of learning is abstraction. Abstraction is the process of removing knowledge from data that has been stored. In order to do this, general notions about the data as a whole must be developed. Application of existing models and Knowledge generation includes both the creation of new models. Training is the process of adjusting a model to a dataset. Once the model has been trained, data is transformed into an abstract form that distils the original information.

3. ***Generalization:***The third stage of learning is generalization. Generalization is the process of converting information about previously saved data into a format that may be used to future action. These procedures must be followed when performing tasks that are comparable to but not identical to those that have already been seen. The objective of generalization is to identify the characteristics of the data that will be most helpful for upcoming activities.

4. ***Evaluation:***Evaluation is the last stage of the learning process. The process of providing feedback to the user allows us to assess the usefulness of the knowledge we have learnt. Then, using this feedback, the entire learning process is improved.

1.4 Machine learning mainly has three types:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement Learning

1. Supervised machine learning

Machine learning that is supervised operates under supervision. It includes a model that can make predictions using data that has been labelled. A tagged dataset is one for which you are already aware of the intended response. The term "labelled data" refers to input data that has already been assigned the appropriate output. The objective of supervised learning is to discover a mapping function that connects the input variable (x) with the output variable [19] algorithm (y).

Supervised learning has applications in the real world such as risk assessment [19], spam filtering, image categorization, fraud detection, etc[19].

2. Unsupervised Learning

A sort of machine learning algorithm called unsupervised learning is used to derive conclusions from datasets that only contain input data without tagged replies. An observation's classification or categorization is not present in unsupervised learning techniques. There is no estimation of functions because there are no output values. Because the learner's examples are unlabeled, it is impossible to assess the accuracy of the structure that the algorithm generates. Cluster analysis, which is used for exploratory data analysis to uncover underlying patterns or grouping in data, is the most used unsupervised learning technique.

3. Reinforcement Learning

Between supervised and uncontrolled learning, this falls. Although the algorithm is informed when the solution is incorrect, it is not informed how to make it right. It must investigate and test out several options before figuring out how to provide the correct response. Because the monitor only assigns a score and makes no suggestions for improvement, reinforcement learning is sometimes referred to as learning with a critic.

The issue of persuading an agent to behave in the real world in a way that maximizes its rewards is known as reinforcement learning. Unlike most forms of machine learning, a learner (the software) is not given instructions on what to do; rather, it must experiment with several options to determine which ones produce the greatest rewards. In the most fascinating and difficult situations, choices can influence not just the immediate reward but also later circumstances and, in turn, all subsequent benefits.

Chapter 2

Background Details and Literature Review

N. Hemageetha*et. al.* "**Radial Basis Function Model for Vegetable Price Prediction**"[2]. In this paper according to researcher developing countries are required more focus on agriculture sector. The price prediction helps the farmers as well as government for the public benefits. This model predicts the price of vegetables. The author used back propagation neural network (BPNN) and Radial basis function neural network (RBF). They used tomato weekly price dataset from January 2009 to march 2012 from www.tnau.ac.in. According to this paper RBF is more efficient and accurate than the back propagation.

Yung-Hsing Peng *et. al.* "**An Investigation of Spacial approaches for Crop Price Forecasting in Different Taiwan Markets**"[3]. In the agribusiness demand and supply indicates the market price. In this paper Nearest Neighbor algorithm, kriging method, Artificial Neural Network and Kriging method is used to get better efficiency. kriging method is good for low error and better efficiency. It takes many factors like climate condition, past price, planting area.

Rohit Kumar Rajak*et. al.***" Crop Recommendation to Maximum Crop Yield using Machine Learning Technique**"[4].according to this paper due to lack of knowledge people not sow the right crop according to yield of land. In this paper author tired to solve this problem using SVM algorithm and ANN to provide high accuracy and efficiency. It use many factors to improve the model like weather condition or climate condition.

Chenhao Wang *et. al.* "**High and Low Prices Prediction of Soybean Futures wit LSTM Neural network** "[5]. It is difficult to predict the future prices. Many researches are done on predicting the closing price. This research is done on the low and high price of soya bean. They used Long short-term memory(LSTM) and Mean Absolute Error in their performance evaluation model. They takes the dataset from Dalian commodity exchange. According to them LSTM is better because it gave 80% more accuracy. To improving its accuracy, it use other attributes also like weather, social conditions etc.

Ashwini Darekar*et. al.* “**Cotton price forecasting in major producing states**”[6]. India is one of the largest cotton producing and exporting country. It is crucial to predict the price of cotton so that not only farmers would profit, but it will also help purchaser and millers too. The goal of this study is, considering the major cotton producing states and forecast the price of cotton. For this study historical data of 10 years, monthly price of cotton, data collected from AGMARKNET. The parameters for the model were estimated using R programming software. Usually, the crops prices are varies depending on global supply and demand along with the domestic demand and supply. So for the further improvement it is required to consider the global demand and supply and also export/import prices.

Zeel Doshi *et. al.* “**AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms**” [7].The majority of the farmers in India decides which crop to grow or sow based on their previous experience and intuition in a particular season. This may affect the yield as well as prices in the market. To solve this issue authors proposed a model is depend on Big Data Analytics and Machine Learning algorithms. In this paper they presented the system called Agro Consultant. The proposed system assists the farmers to make decision about which crop to grow based on geographical location, environmental factors and soil characteristics. The model presented in this paper can be improved by considering the crop rotations. That is which crop to grow in the current season based on the crop was harvested in the previous cycle or season.

Lakshmi. N *et. al.* “**Crop Recommendation System for Precision Agriculture**” **[8**]. Big Data is one of the useful technology for developing agriculture, medicines and retails recommendation system. While developing the new model Big Data, Data mining and IOT play an important role. In the current research paper the authors proposed a model which predict the types of soil and crop that is suitable to be grown in that soil. They used flume to download the data from agricultural website and saved them on Hadoop file system. The data processed by Map Reduce. The proposed model helps farmers to sow the right seed based on the soil and whether condition. This model can be further improved with the data set containing more number of attributes in yield prediction.

Chenhao Wang *et. al.* "**High and Low Prices Prediction of Soybean Futures with LSTM Neural Network**" [28]. The prediction of future prices is a challenging task. Most of the research done previously concentrated on predicting the closing price. But it is important to predict low and high prices too. Here the authors came up with a model for predicting the high and low prices of soybean. They used LSTM for developing the proposed model. They have used the dataset from the Dalian Commodity Exchange. Later they used mean absolute error (MAE) and trend accuracy to evaluate the performance of the model [20]. With this study they claimed that LSTM performance is better. In trend estimation it gave more than 80% of the accuracy. The authors used only the trading price or values considered. As a further improving the it is required to use other related attributes like weather, social conditions etc.

Yongli Zhang *et. al.* "**A Novel Agricultural Commodity Price Forecasting Model Based on Fuzzy Information Granulation and MEA-SVM Model**"[9].In the agriculture few important things are reduce the market risk. Also, it is required to increasing the agricultural income. Sometime government regulation influences the price of commodities. Considering all these it is important to predict the agricultural commodities price. With this paper the authors came up with a forecasting model for agriculture commodities price. To build the proposed model the fuzzy information granulation[21], MEA algorithm and support vector machine[21] are used. In this study the authors claimed that the MEA-SVM model was effective and[21] with this we can get higher accuracy during prediction. Also calculation speed during the forecasting is high in case of MEA-SVM model. There is scope for improving the proposed model by reducing the long-term prediction error.

A.Amarender Reddy *et. al.* "**Price Forecasting of Tomatoes**"[10].The price fluctuations of commodities based on supply and demand. Tomato is one of the perishable commodities and it is important to forecast harvest period tomato prices. So that farmers can take informed production decisions. In other models seasonal variations are not considered. This model is based on seasonal variations of tomato prices from January 2006 to December 2016. For forecasting the prices It was done using the SARIMA model. This study shows that there is a difference in tomato prices across different regions due to lack of cold storage and refrigerated transport facility.

Tao Xiong*et. al.* “**Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China**”[11]. Previously the limited attention paid for the research. Here authors studied and proposed a hybrid method. They combined trend decomposition methods. This model is proposed based on Extreme Learning Machines and STL. Using this model authors tried to forecast vegetable prices for short, medium and long term. The authors mentioned that SARIMA is very popular forecasting technique. In case of SARIMA the performance may be poor, if we assume time series with high nonlinearity. Further this can be improved by interval forecasting of vegetable prices.

Koki Yamamoto *et. al.* “**Predicting System of Harvest Time And Yield of Tomato**”[12].Even farmers produce thousands of tons of tomato, many time farmers have to sell the same for lower price. This is demand and supply issue. On this research the authors proposed a new system. Based on the temperature these proposed models calculate harvest days and yield. With this system farmers can take decision on how to produce to get better profit. The limitations of this system is , only the temperature attribute is used for forecasting and it is good to consider other parameters too for getting better performance or accuracy.

Gerasimos Rigatos *et. al.* “**Forecasting of commodities prices using a multi-factor PDE model and Kalman filtering**”[13].In this study the authors proposed a method for forecasting commodities prices. For this purpose, they used Kalman filtering and Schwartz partial differential equation. The PDE model is a well-liked model for depicting changes in the price of agricultural commodities. This is model is very useful especially in case of long-term contracts. As an enhancement, to get more accurate prediction of commodities price as an m-step predictor, the Kalman filter can be created.

Yuchen Weng *et. al.* “**Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler**”[14]. Agriculture-related goods have a significant role in the supply chain. The prices of commodities affected by many factors such as climate condition, government policies, supply and demand etc. The forecasting the prices of horticulture products are important because we can’t store them for long time. In this paper the authors came up with a model to forecast the prices for agriculture products using

Back Propagation network, Auto-Regressive Integrated Moving Average model, and Recurrent Neural Network method. In the future this model can be improved by considering the social factors (policies, holidays), market condition, and weather condition. For the generalization it is good to consider the price data of different regions. Also, parallel learning is the another option can be considered to train the computation system.

Asma HasifaNurcahyono*et. al.* “**Price Prediction of Chili in Bandung Regency Using Support Vector Machine (SVM) Optimized with an Adaptive Neuro-Fuzzy Inference System (ANFIS)”**[15]. Indonesia is the largest chili producing country. In this country the farmers facing the problem of price fluctuations. One of the main reasons is weather, because chili plants are sensitive and damaged if exposed to too much water/raining [21]. In this paper researchers had done a research to predict the chili price. They used SVM algorithm, long with Adaptive Neuro Fuzzy Inference System [21]. The authors claimed that with SVM algorithm optimized with Adaptive Neuro-Fuzzy Inference System [21] we can get better performance and accuracy compared to Support Vector Machine algorithmalone. As a further improvement it is good to consider other conditions like soil, rain, temperature etc.

Dabin Zhang *et. al.* “**Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons**”[16]. The fluctuations of agricultural products has great impact on both general public and farmer. So it is essential to have a system to forecast the commodity process which helps to take decisions. In this research paper the authors discussed about a forecasting model developed using Support Vector Regression, Extreme Learning Machine techniques and Artificial Neural Network. To measure or to learn the performance of the proposed model Support Vector Machine techniques and Random Forest are applied. Maximum Relevance Approach is used to improve the forecast accuracy. The experimental [21] result shows that proposal model has better performance when the simple model average and optimal candidate model are used. This model can be further improved by using some powerful classifiers like AdaBoost and Bayesian networks.

The Table 2.1 below shows the comparison of the existing models and their pros and cons. Also scope for future work is listed.

Sr.No	Author	Title	Year	Methodology /Framework	Application	Limitations/ Future scope
1.	N. Hemageetha	Reference[2]	2013	Back-propagation neural networks and Radial basis function neural networks.	Prediction of tomato price.	To increase the prediction accuracy genetic algorithm based neural network will be good choice. Only 5 year historical price considered.
2.	Yung-Hsing Peng	Reference[3]	2015	Nearest Neighbour. The inverse distance weighting. The Kriging method, and the Artificial Neural Network	Crop price forecasting	Data set with large number of attributes to be considered, like climate, region etc. Implement yield prediction.
3.	Rohit Kumar Rajak	Reference[4]	2017	Majority voting technique using Artificial Neural Network and Support Vector Machine.	Crop recommendation based on site specific parameters.	Only Soil specific attributes used.
4.	Chenhao Wang	Reference[5]	2018	LSTM neural network, Mean Absolute Error (MAE).	Instead of only closing price, this model predicts highand low prices.	Only trading price or values considered.
5.	Ashwini Darekar	Reference [6]	2017	ARIMA	Future prices Prediction of cotton.	The proposed technique does not guarantee perfect forecasts.

6.	Zeel Doshi	Reference [7]	2018	Random Forest ,Linear Regression, Decision Tree, KNN, and Neural Network frameworks used.	Which crop to grow based on soil. Rain fall predictor.	Consider crop rotations. Consider other aspects like demand supply, harvest prices and retail prices.
7.	Lakshmi. N	Reference[8]	2018	Big Data is used. Flume is used for download information.	Predict the types of soil and the crops that are suitable to begrown in that soil.	Data set with a greater number of attributes like weather, geographical features, water utility to be considered in yield prediction.
8.	Yongli Zhang	Reference [9]	2018	Information granulation by Fuzzy system, MEA, and SVM are used.	Agricultural commodity price forecasting.	This model has larger error for the long-term prediction. The current model can be further improved by optimizing the SVM model's kernel parameter g and penalty parameter c.
9.	A. Amarender Reddy	Reference[10]	2018	SARIMA	Price forecasting of Tomato.	Social factors not considered.
10.	Tao Xiong	Reference[11]	2018	Loess (STL) Extreme learning machine (ELM) Hybrid STL-ELM method	Price forecasting for vegetables.	The limitation of this model is attention given only to point forecasting. The interval forecasting of prices is required.
11.	Koki Yamamoto	Reference[12]	2018	Temperature sensors and RNN	This model calculates harvest days and yield of tomatoes. Also predict the harvest time.	Only Temperature is considered.

12.	Gerasimos Rigatos	Reference[13]	2018	Schwartz PDE and Kalman filtering.	Forecasting commodities prices	To get the more accuracy in estimation, redesigning the Kalman filter as a m-step ahead predictor is required.
13.	Yuchen Weng	Reference[14]	2019	ARIMA, Back Propagation network method, and RNN. Data collection - web crawler technology	Price forecasting	By considering the social factors like, holidays, festivals, polices etc. it is possible to improve RNN model. While training the computation system, parallel learning can be applied.
14.	Asma Hasifa Nurcahyono et al.	Reference[15]	2019	SVM, Optimized Adaptive Neuro- Fuzzy Inference System.	Price Prediction of Chili	Only 3 years weather data used.
15.	Dabin Zhang	Reference[16]	2020	Extreme Learning Machine, Support Vector Regression, and Artificial Neural Network. The approach of "Minimum Redundancy and Maximum Relevance" was applied.	Forecasting agricultural commodity prices	To improve the classification capability, AdaBoost and Bayesian networks could be used.

Table 2.1 Comparison between the research papers

The literature offers a variety of statistical and intelligent forecasting techniques for the prices of agricultural commodities.

The "no free lunch" theory [17] states that there is no one model that works for all commodities. It can be challenging for people to decide which model is best when dealing with a fresh class of agricultural product for this particular forecasting activity. Of course, decision-makers can assess the effectiveness of a number of widely used forecasting approaches and select the most advantageous one. However, it takes a lot of time to train many models. It goes without saying that a quick and automatic algorithm is required to choose the best forecasting technique for agricultural commodities.

2.1 Objectives

1. The main aim of the project is to predict the crop price with the machine learning algorithms on demand in the market from the past.
2. It will help the government to take efficient steps towards the farmers.
3. It will help to reduce the suicide rates.

Chapter 3

Design or Framework of the project work

3.1 Programming Language:

3.1.1 Python

Python is a flexible, high-level, dynamic, and interpreted computer language. It allows the development of programs using an object-oriented approach. It offers a large number of high-level data structures and is straightforward and simple to learn. Because it is straightforward to learn but robust and flexible, Python is a programming language that is appealing for the development of applications. Python is the best language for scripting and quick application development because of its syntax, dynamic typing, and nature as an interpreted language. Python supports a variety of programming patterns, including imperative, functional, and object-oriented programming patterns. Python is not intended for usage in certain contexts, such as web development. It is referred to due to the fact that it can be utilized with online, enterprise, 3D CAD, etc. as a multipurpose programming language. [23]. Since the variables are dynamically typed, we don't need to [23] identify their data types because To give an integer variable the value 10, we can just type a=10. Python requires no compilation, which expedites development and debugging [22] stage required and the edit-test-debug cycle is quite short.

3.1.2 Pandas Package:

An open-source Python library that provides high-performance data manipulation is referred to as Pandas. An Econometrics from Multidimensional Data is known as Pandas., gets its name from the phrase panel data. Wes McKinney created it in 2008 and uses Python to analyze data.

Data analysis requires a number of processing processes, including combining, cleaning, and restructuring. For quick data processing, there are a variety of tools available, including NumPy, SciPy, Cython, and Panda. However, we favor Pandas since they are quicker, easier, and more expressive to utilize than other tools.

Due to the fact that Pandas is built on top of the NumPy package, NumPy is necessary in order to use Pandas.

Prior to Pandas, Python could manage data preparation, but it only provided a limited set of capabilities for data analysis. When Pandas arrived, data analysis's capabilities were enhanced. No matter where the data originated from, it can do the five essential steps—load, edit, prepare, model, and analyze—that are required for data processing and analysis.

3.1.3 NumPy Package:

Numeric Python is a Python module for computing and processing multidimensional and linear array elements; it is also referred to as NumPy.

Travis Oliphant developed the NumPy package in 2005 by integrating the capabilities of the ancestor module Numeric into another module named Numarray.

It is a C-based extension module for the Python programming language. It offers a variety of routines that can perform numerical calculations quickly.

NumPy offers a wide range of reliable data structures, such as multidimensional arrays and matrices. These data structures are used to do the best calculations on arrays and matrices.

3.1.4 Sklearn package:

Skearn is the name of the most trustworthy and powerful Python machine learning package (Skit-Learn). It provides various efficient[21] techniques for statistical modelling and machine learning, including as classification, regression, clustering, and dimensionality reduction, through a Python consistency interface. This library was primarily [21] created in Python and is built on NumPy, SciPy, and Matplotlib.

3.1.5 Matplotlib package:

Matplotlib is one of the most used Python data visualization tools. It is a multi-platform package that creates 2D charts from arrays of data. NumPy, Python's extension for numerical mathematics, is used by Matplotlib, which is written in Python. Plot integration with programmes that employ Python GUI toolkits like PyQt and WxPythonotTkinter is facilitated by its object-oriented API. Additionally, it works with web application servers, IPython and Python shells, and Jupyter notebooks.

The proprietary programming language created by MathWorks, MATLAB, offers a procedural interface called Pylab that is part of Matplotlib. It is possible to think of Matplotlib and NumPy as the open-source alternatives to MATLAB.

John D. Hunter created Matplotlib for the first time in 2003. The most recent stable version, 2.2.0, was made available in January 2018.

3.2 DESIGN

The design system for the crop price prediction model, which is impacted by demand and price, is represented in this section.

3.2.1 Dataset

The report's data on paddy prices was obtained from the Indian government's agriculture website, www.data.gov.in. [24]. From 1985 to 2021, the dataset includes data on production, demand, crop year, season, area, and yield. Price is taken as MSP(minimum support price) which is fixed by the government on many determinates.

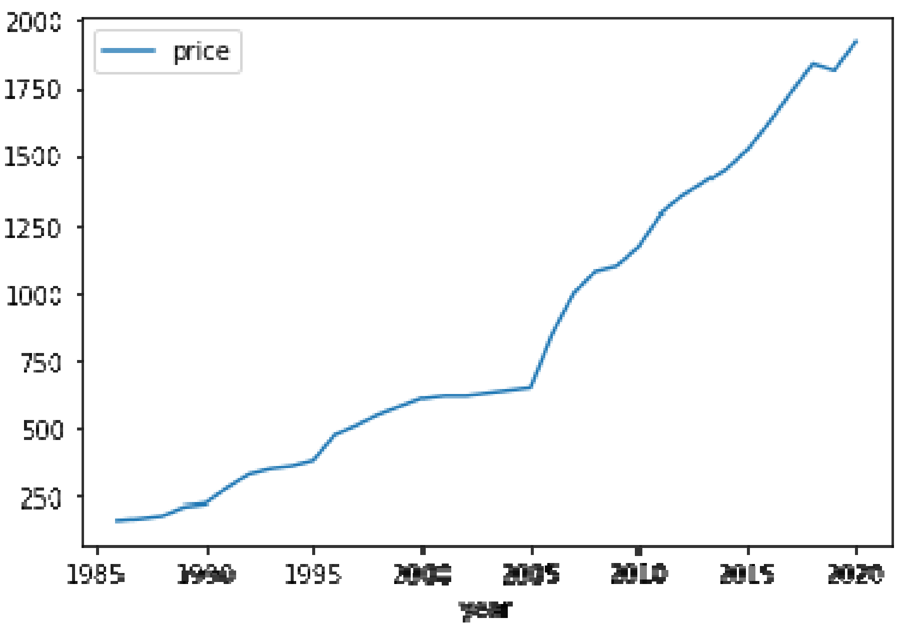


Figure 3.1: Comparison between year and price

The figure 3.1 represents the price of wheat with comparsion of years as we can see in the graph the price increases continuously this is only happened due to increase in production. The data is collected from the official sites and github which is used in previous research paper. We taken the dataset from the github repository [26].

3.2.2 MSP Determinants

The Commission maintains in mind the numerous Terms of Reference (ToR) granted to CACP in 2009 when recommending price policies for various commodities under its purview. As a result, it analyses

1. Supply and demand
2. Production costs
3. Market pricing movements, both domestically and internationally
4. Price parity between crops
5. Agriculture and non-agriculture trade phrases
6. A minimum margin of 50% above the cost of manufacture
7. Potential effects of MSP on the product's customers.

It should be highlighted that while manufacturing costs play a significant role in determining MSP, they by no means serve as the sole determinant of MSP.

3.2.3 Demand - What Is It?

Demand in economics refers to both a consumer's willingness to pay a price for a [25] certain good or service and their desire to purchase such products. All other factors being equal, as an item or service's price rises, so does the amount requested. The total amount sought after by all buyers in a market for a certain product is known as the market demand. The total quantity of demand for all commodities and services in an economy is known as aggregate demand.

The following five factors influence demand:

1. The cost of the product or service
2. The buyers' income
3. The costs of complementary or alternative goods or services that are acquired with a specific item or as a replacement for a product.
4. Consumer preferences or tastes will determine demand
5. Consumer expectations regarding whether product costs will increase or decrease in the future

Factors affecting the demand:

1. The product's pricing
2. The quantity of manufacturers
3. The input costs
4. The price of other possible products
5. Unpredictable factors such as weather.

3.2.4 Supply: What Is It?

The total amount of a specific [25] commodity or service that a provider makes accessible to clients at a specific time and price is referred to as supply in economics. Usually, it is decided by market action. For instance, a supplier might raise their supply in response to rising demand.

Determinants of Supply

1. Price
2. The number of sellers in the market
3. The price of resources used to produce the product
4. Tax rates and subsidies
5. Improvements in technology and automation
6. Expectations of the suppliers
7. The price of related products

Factors affecting supply of agricultural products

1. Time
2. Producers' cost structures
3. Predictions for producer prices
4. Product storage capacity
5. How simple it is to switch from one product's production to another

CHAPTER 4

Discussion and Analysis of Results

4.1 Describe regression.

Regression's major objective is to build an effective model that can predict the dependent attributes from a variety of attribute variables. When the output variable, such as a salary, weight, or area, is real or continuous, there is a regression problem.

Linear Regression: What Is It?

A sort of regression in which there is a linear relationship between the independent and dependent variables is known as simple linear regression. The diagram's straight line is the line that best fits the data. Simple linear regression seeks to fit the model as closely as feasible by taking into account the available data points and presenting the best fit line.

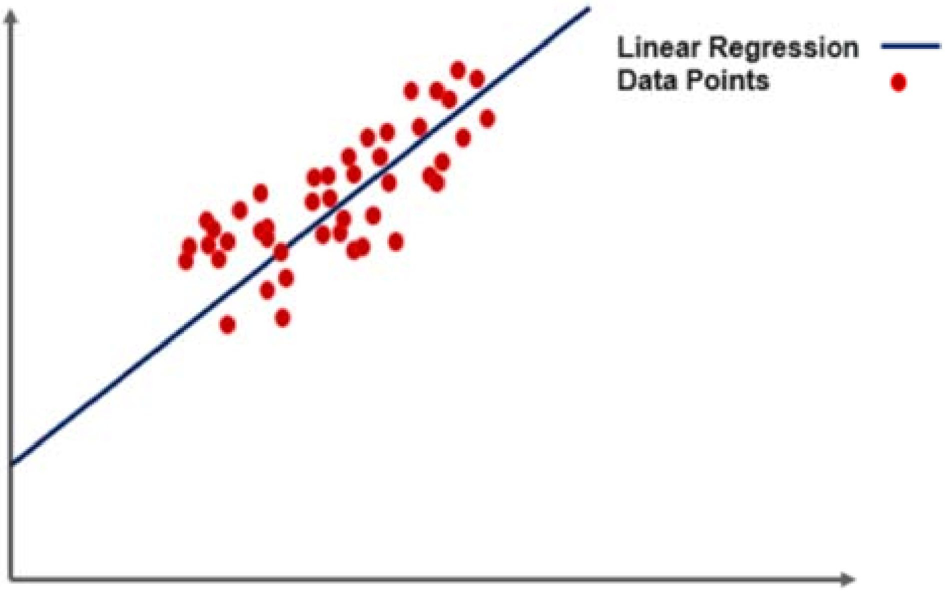


Figure 4.1.1 Linear Regression

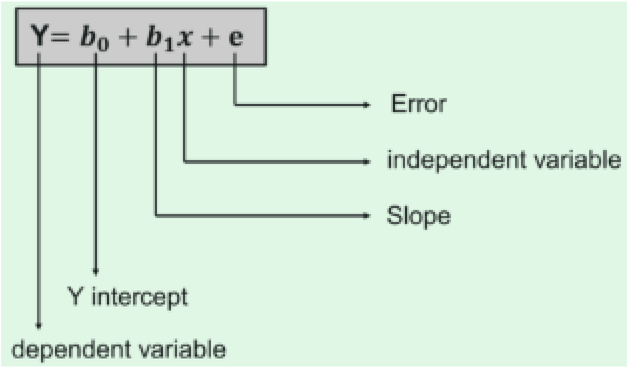
Let's look at a few key terms in simple linear regression before moving on to the linear regression algorithm's operation.

Terminologies for Linear Regression

Before tackling the linear regression process, it's critical to understand the following jargon.

Function Cost

The linear equation shown below can be used to determine the best fit line.



….(1)

1. Y stands for the dependent variable that needs to be forecasted.
2. The intercept b0 designates the point on a line where it touches the y-axis.
3. X stands for the independent factors that affect the prediction of Y, and b1 is the line's slope.
4. The letter e stands for the mistake in the final prediction.

The cost function provides the ideal values for b0 and b1 to produce the line that fits the data points the best. By converting this issue into a minimization issue, the ideal values for b0 and b1 are discovered. The difference between the anticipated value and the actual value in this case is kept to a minimum.

$$\text{minimize} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{pred_i} - \mathrm{y_i})^2$$

$$\mathrm{J} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{pred_i} - \mathrm{y_i})^2 \text{....(2)}$$

Gradient Descent

Gradient descent is the next crucial phrase to comprehend while discussing linear regression. It is a technique for altering the values of b0 and b1 to lower the MSE[24]. The goal is to iterate the b0 and b1 values continuously until the MSE is minimized.

We use the gradients from the cost function to update b0 and b1[25]. We take partial derivatives with regard to b0 and b1 to determine these gradients. The gradients are the partial derivatives that are utilized to update the values of b0 and b1[24].

Uses of linear regression

1. Housing applications to forecast prices and other elements;
2. Risk analysis;
3. Finance applications to forecast stock prices, investment analysis, etc.

Model Evaluation

```
[12] from sklearn import metrics
     print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(y_test,preds))
     print("Mean Squared Error (MSE): ",metrics.mean_squared_error(y_test,preds))
     print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(y_test,preds)))

Mean Absolute Error (MAE):  61.82403000479639
Mean Squared Error (MSE):  7324.037383365062
Root Mean Squared Error (RMSE):  85.58058999191968
```

Figure 4.1.2 linear regression modal evaluation result

After the process we get above the result which represents in the figure 4.1.2.

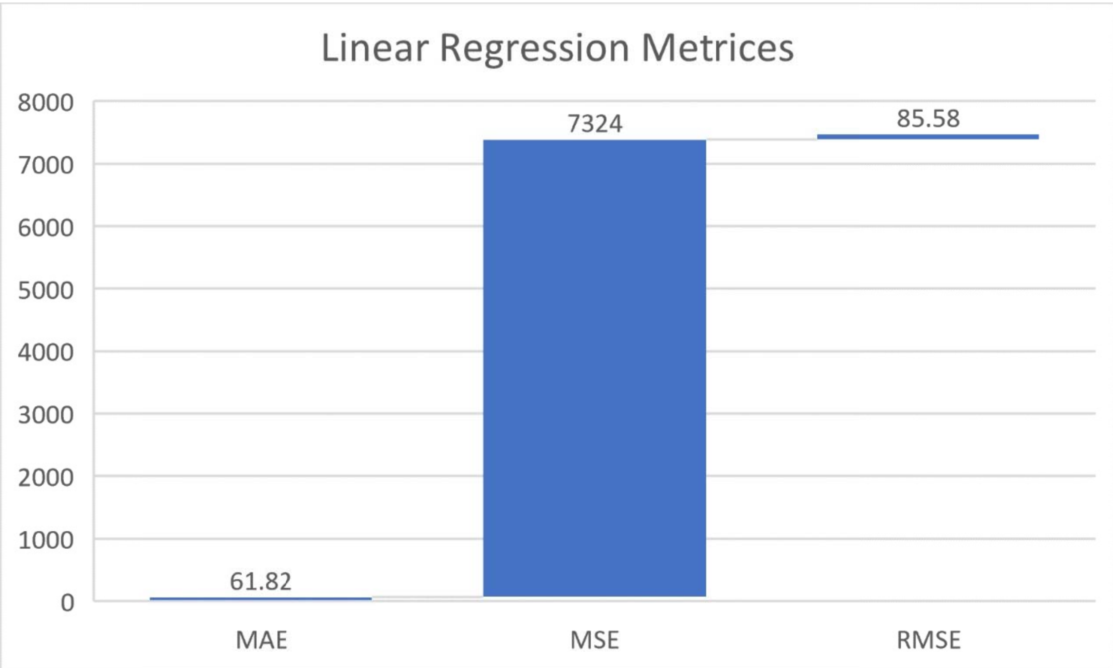


Figure 4.1.3 Linear Regression Metrices

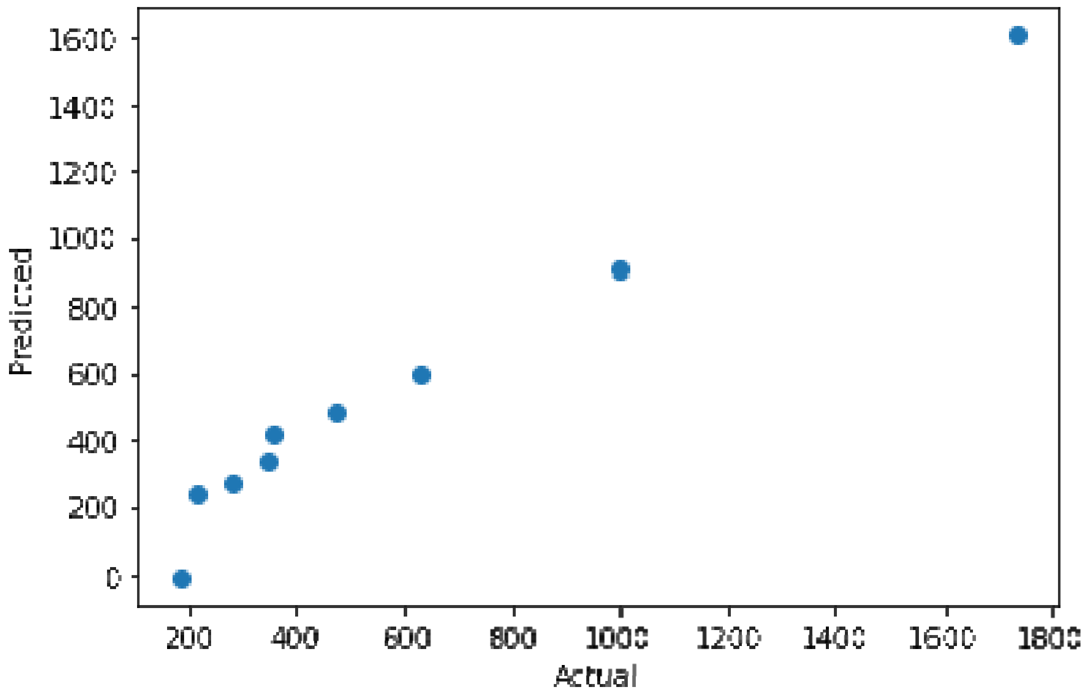


Figure 4.1.4 output with actual and predicted price

The above fig shows the difference between the actual price and the predicted price of the crop. The accuracy of the model is 96.71%.

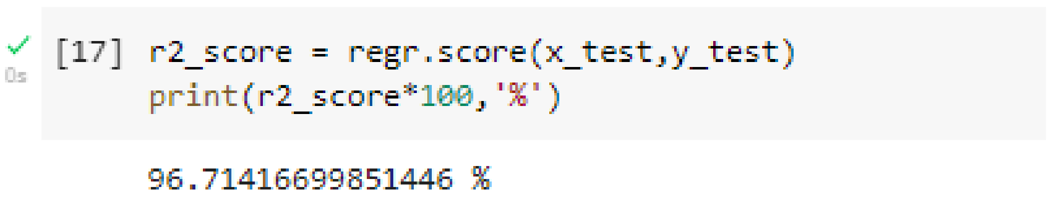


Figure 4.1.5 accuracy of the modal

4.2 Multiple Regression

One of the key regression techniques, multiple linear regression simulates the linear relationship between a single[22] continuous dependent variable and a number of independent variables.

The independent or predictor variable may be continuous or categorical for MLR, but the dependent or target variable (Y) must be continuous[22].

Each feature variable needs to simulate the dependent variable's linear relationship.

A regression line is attempted to be fitted using MLR through a multidimensional space of data points.

$$Y = \mathbf{b_0} + \mathbf{b_1 x_1} + \mathbf{b_2 x_2} + \ldots\ldots + \mathbf{b_n x_n} \ldots\ldots (4.2.1)$$

Where

Y is the output or response variable and

b0, b1, b2, b3, bn, etc. are the model's coefficients.

Various Independent/feature variables, x1, x2, x3, x4,...

```
farming_data.plot("year","price")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fda288c8910>
```

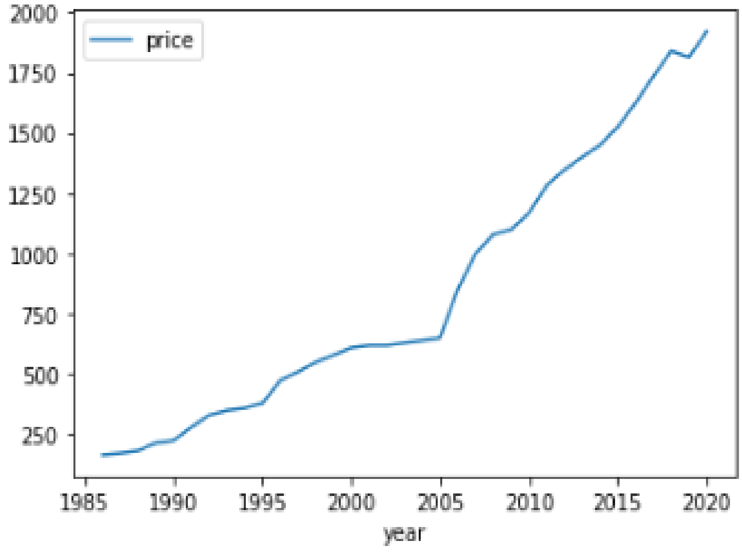


Figure 4.2.1 graph between the price and year

This graph represents the relation between the price and year with the help of linear regression which is further used for residuals and to find mean squared error.

```
[12] from sklearn import metrics

     print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(y_test,preds))
     print("Mean Squared Error (MSE): ",metrics.mean_squared_error(y_test,preds))
     print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(y_test,preds)))

     Mean Absolute Error (MAE):  149.89907474909455
     Mean Squared Error (MSE):  33559.135344092334
     Root Mean Squared Error (RMSE):  183.1915263981725
```

Figure 4.2.2 multiple regression modal evaluation result

After the process we get above the result which represents in the figure 4.2.2.

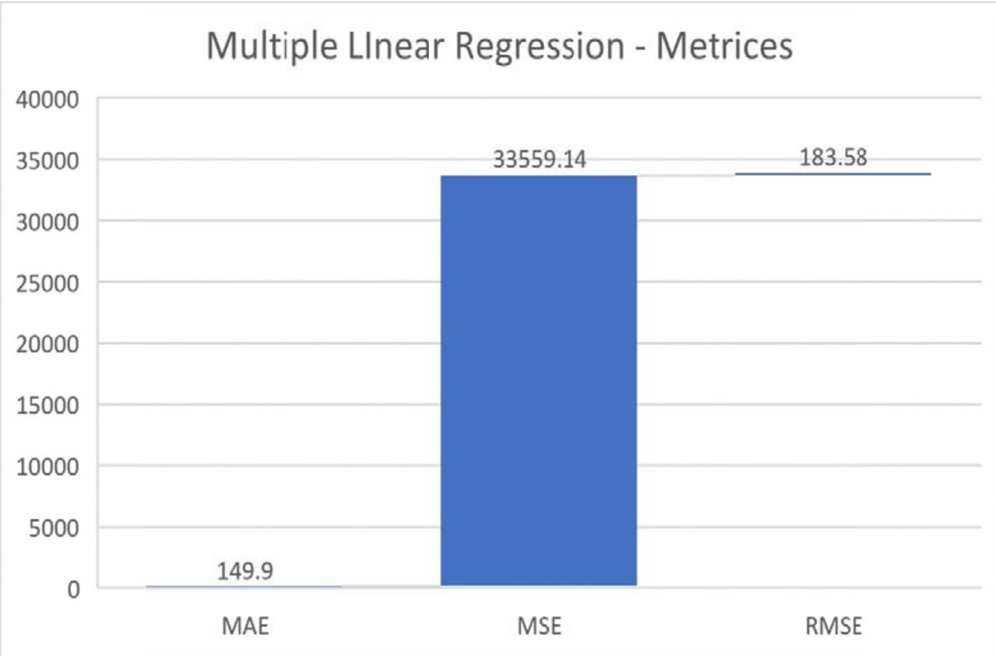


Figure 4.2.3 Multiple Linear Regression Metrices

```
plt.hist(residuals)

(array([1., 1., 1., 2., 0., 0., 0., 1., 0., 1.]),
 array([-232.86587172, -176.88442   , -120.90296827,  -64.92151655,
          -8.94006483,   47.0413869 ,  103.02283862,  159.00429035,
         214.98574207,  270.96719379,  326.94864552]),
 <a list of 10 Patch objects>)
```

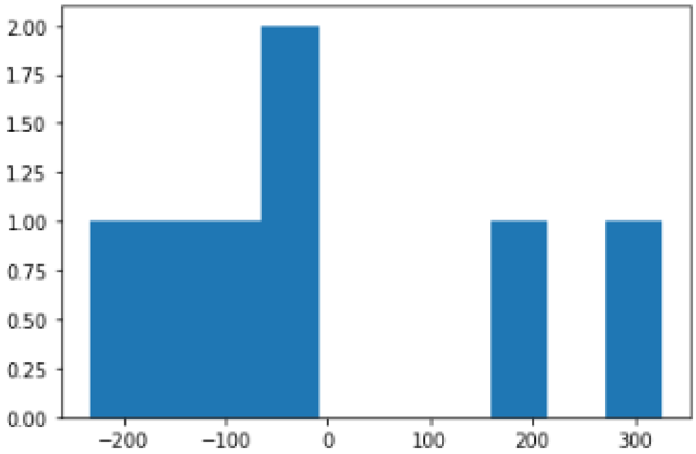


Figure 4.2.4 shows the graph of residuals and mean_squared_error

After the process we get above the result which represents the graph of residuals.

```
[19] plt.scatter(y_test,preds);
     plt.xlabel('Actual');
     plt.ylabel('Predicted');
```

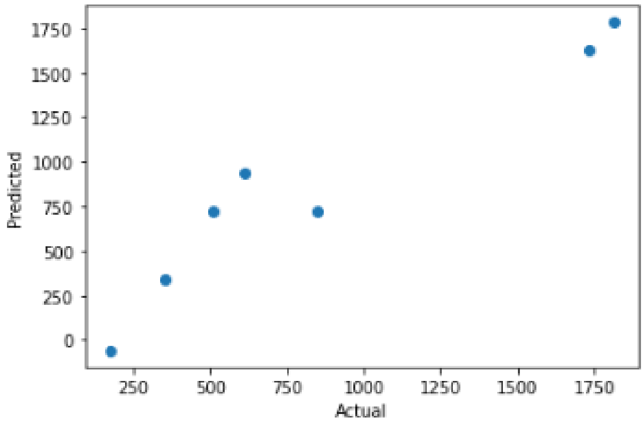


Figure 4.2.5 comparison between the actual and predicated price

The above fig shows the difference between the actual price and the predicted price of the crop. The accuracy of the model is 90.94%.

```
[18] r2_score = regr.score(x_test,y_test)
     print(r2_score*100,'%')

     90.94654427786791 %
```

Figure 4.2.6 shows the accuracy of the modal

4.3 KNN

The k-nearest neighbours (KNN) technique calculates the likelihood that a data point will belong to one group or another based on which group the data points closest to it do[22].

An example of a supervised machine learning[22] technique used to resolve classification and regression issues is the k-nearest neighbour algorithm. However, classification issues are its primary application.KNN is a lazy learning and non-parametric algorithm.

Because it doesn't perform any training when you[21] provide the training data, it is known as a lazy learning algorithm or lazy learner. It doesn't make any calculations during the training period, just keeps the data. Before a query is run on the dataset, a model is not built. KNN is hence perfect for data mining.Because it doesn't make any assumptions about the distribution of the underlying data, it is regarded as a non-parametric method. KNN, in short, seeks to identify the group to which a data point belongs by examining the data points around it.

KNN classifier

```
[14] from sklearn import metrics

     print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(y_test,pred))
     print("Mean Squared Error (MSE): ",metrics.mean_squared_error(y_test,pred))
     print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(y_test,pred)))

     Mean Absolute Error (MAE):  182.42857142857142
     Mean Squared Error (MSE):  79012.71428571429
     Root Mean Squared Error (RMSE):  281.09200324042354
```

Figure 4.3.1 KNN classifier evaluation result

After the process we get above the result which represents in the figure 4.3.1

```
[13] print(knn.score(X_train, y_train)*100,'%')

     39.285714285714285 %
```

Figure 4.3.2 accuracy of the model

Figure 4.3.2 shows the accuracy of the model is 39.28%.

```
plt.show()

No handles with labels found to put in legend.
```

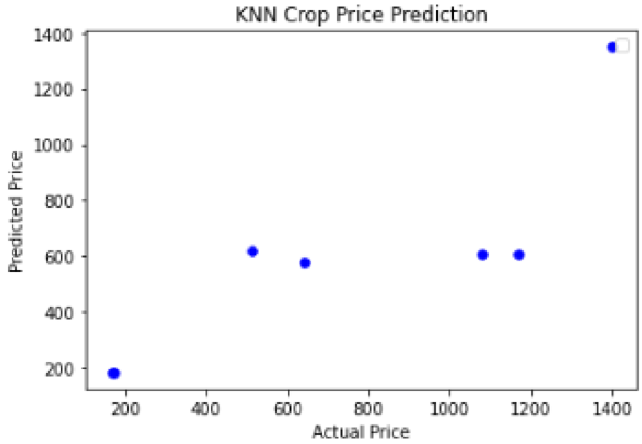


Figure 4.3.3 comparison between the actual and predicated price

The above fig shows the difference between the actual price and the predicted price of the crop.

KNN Regression

```
# function to show the plot
plt.show()
```

```
No handles with labels found to put in legend.
```

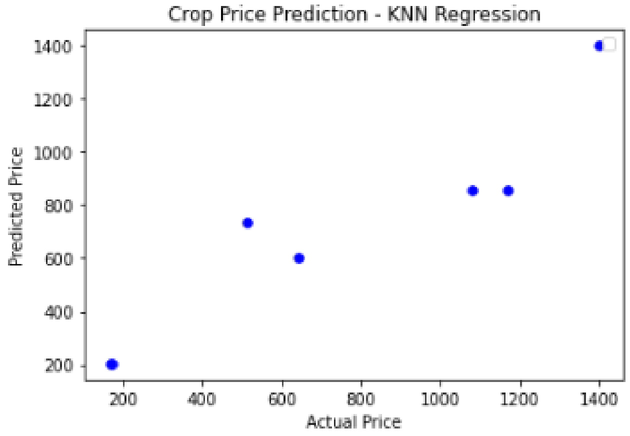


Figure 4.3.3 comparison between the actual and predicated price

The above fig shows the difference between the actual price and the predicted price of the crop.

```
[13] print(knn.score(X_train, y_train)*100,'%')

96.94184142541444 %
```

Figure 4.3.4 accuracy of the model

Figure 4.3.4 shows the accuracy of the model is 96.94%.

```
[14] from sklearn import metrics
     print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(y_test,pred))
     print("Mean Squared Error (MSE): ",metrics.mean_squared_error(y_test,pred))
     print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(y_test,pred)))

Mean Absolute Error (MAE):  124.85714285714286
Mean Squared Error (MSE):  29211.428571428572
Root Mean Squared Error (RMSE):  170.9135119627134
```

Figure 4.3.5 KNN evaluation result

After the process we get above the result which represents in the figure 4.3.5.

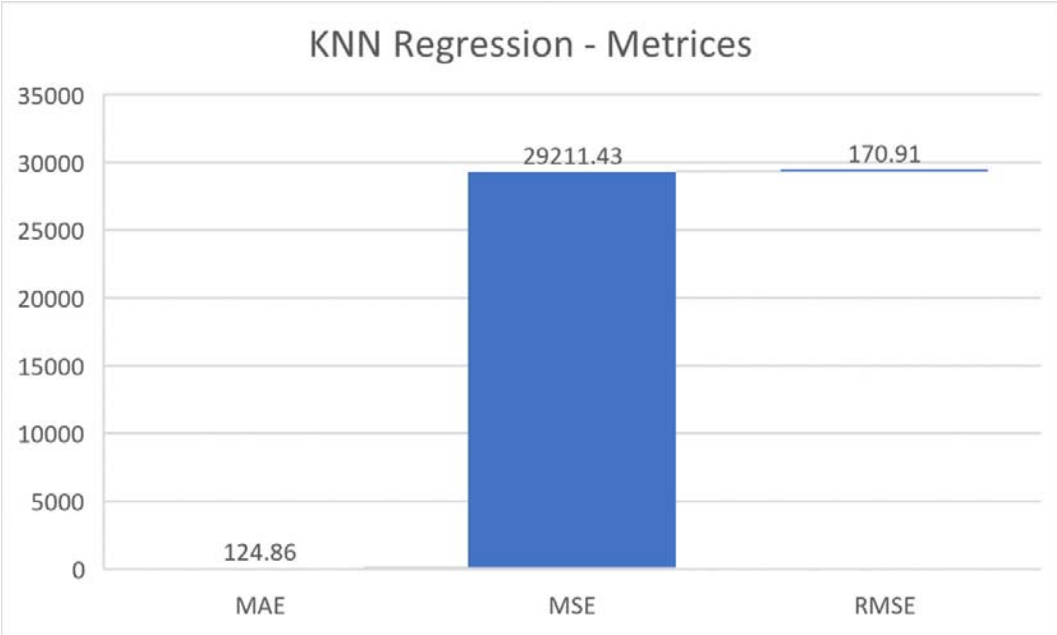


Figure 4.3.6 KNN Regression Metrices

4.4 Gradient Boosting Regression

"Boosting" is a technique for fusing various straightforward models into a single composite model. Due to the addition of single simple models (also known as weak learners) while maintaining the integrity of the model's current trees, boosting is also known as an additive model. The final, complete model improves as we merge more and more simple models, making it a better predictor. Since the process employs gradient descent to minimize the loss, the name "gradient" in "gradient boosting" refers to this technique.

Gradient boost is utilized for regression when to forecast a continuous value, it is employed., such as age, weight, or cost. This differs from applying linear regression. We'll stick to regression in this tutorial because this is a little different from the configuration used for classification.

In gradient boosting, decision trees are utilized as the weak learners.To address the machine learning issue, Decision Tree presents the data as a tree. The interior nodes of the tree representation stand for attributes, whereas each leaf node represents a class label. The squared error is often the loss function (particularly for regression problems). There must be differentiability in the loss function.

Gradient Boosting Regression uses the same residuals concepts as linear regression.

Relative describes this variation. Then, using gradient boosting regression, a weak model that associates features with that residual is trained. This procedure moves the model closer to the desired outcome by adding the residual predicted by a weak model to the input of the current model. This process can be repeated numerous times to enhance the overall model forecast.

It should be mentioned that although Gradient Boosting Classification is used to predict classes like whether a patient has a specific disease or not.

The following are the general steps we take to develop Gradient Boosting Regression:

1. Choose a poor learner,
2. apply an additive model,
3. specify a loss function,
4. reduce the loss function.

Gains from Gradient Boosting

1. improved precision
2. Reduced pre-processing
3. greater adaptability
4. Lack of data

```
x_ax = range(len(ytest))
plt.plot(x_ax, ytest, lw=0.8, color="blue", label="original")
plt.plot(x_ax, ypred, lw=0.8, color="red", label="predicted")
plt.legend()
plt.show()
```

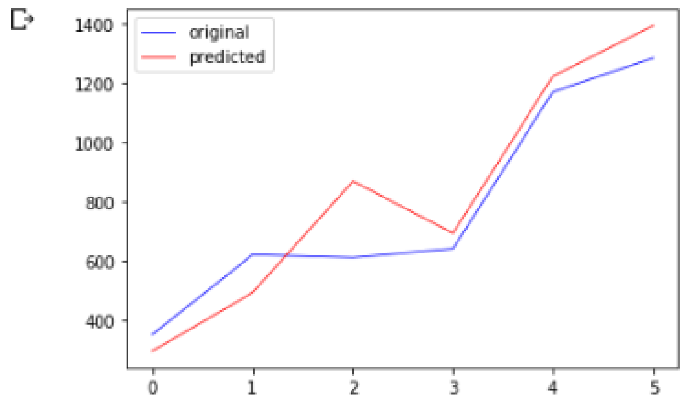


Figure 4.4.1 comparison between the actual and predicated price

The output from the Gradient Boosting Regression is shown in the abovefigure 4.4.1 with the help of graph.

```
#predict accuracy
print(gbr.score(xtest,ytest)*100,"%")
```

```
84.41049645800692 %
```

Figure 4.4.2 accuracy of the model

Figure 4.4.2 shows the accuracy of the model is 84.41%.

```
ypred = gbr.predict(xtest)
import numpy as np

from sklearn import metrics
print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(ytest,ypred))
print("Mean Squared Error (MSE): ",metrics.mean_squared_error(ytest,ypred))
print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(ytest,ypred)))
print(gbr.score(xtest,ytest)*100,"%")
```

```
Mean Absolute Error (MAE):  109.59519321357988
Mean Squared Error (MSE):  17307.48848440897
Root Mean Squared Error (RMSE):  131.5579282461113
84.41049645800692 %
```

Figure 4.4.3 Gradient Boosting Regression **modal evaluation result**

After the process we get above the result which represents in the figure 4.4.2.

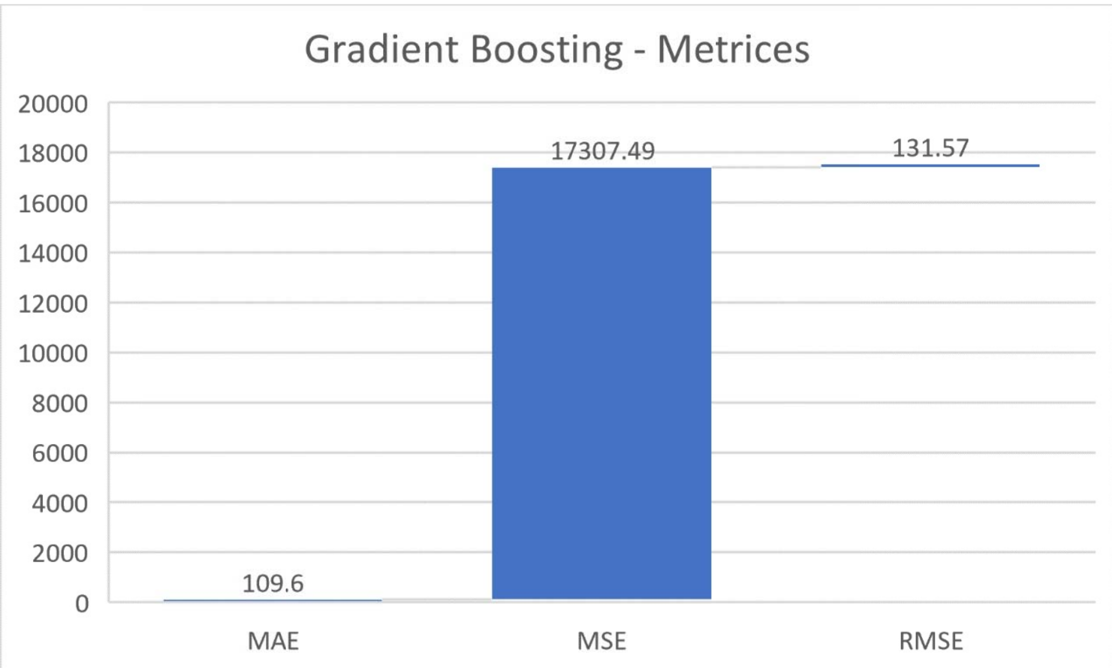


Figure 4.4.4 Gradient Boosting Regression Metrices

4.5 Decision Tree Regresssion

It is a tool with applications in numerous industries. Classification and regression issues can be handled with decision trees. The term itself implies that it employs a flowchart to describe the predictions that come from a sequence of feature-based splits, simulating a tree structure. The leaves at the end, which come after the root node, decide.

some of the terminologies.

Root Nodes – A decision tree's root node is the node from which the population first begins to branch out based on different features.

Decision Nodes – These are the nodes that are produced when the root nodes are split apart.

Leaf Nodes - The nodes that forbid further splitting are known as leaf nodes or terminal nodes.***Sub-tree*** – Similar to how a small area of a graph is referred to as a sub-graph, this decision tree's sub-section is known as a sub-tree.

Pruning – Pruning simply involves removing certain nodes to prevent overfitting.

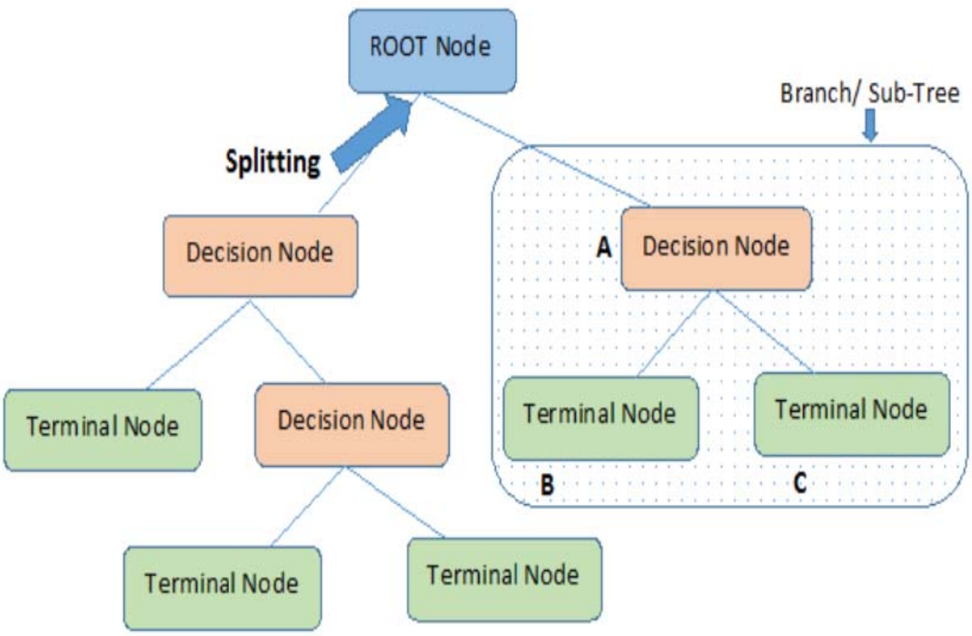


Figure 4.5.1 Representation of Decision Tree

Entropy

Entropy is the term we use to describe the degree of uncertainty in our dataset or measure of disorder.

The following is a formula for entropy:

$$E(S) = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)} \ldots..(1)$$

Here,

p+ stands for the likelihood of a positive class.

The probability of a negative class is p-.

Information Gain

Information gain determines which attribute should be chosen as a decision node or root node by measuring the amount of uncertainty reduced by a given feature.

$$\text{Information Gain} = \mathrm{E(Y)} - E(Y|x) \qquad \ldots.(2)$$

It is simply the entropy of the entire dataset, adjusted for some attribute.

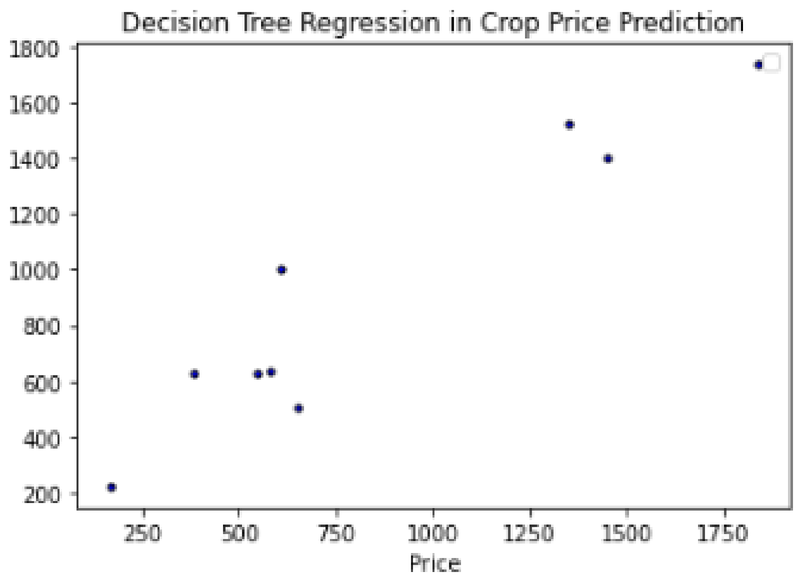


Figure 4.5.1 comparison between the actual and predicated price

The output from the Decision treeis shown in the abovefigure 4.5.1 with the help of graph.

```
[10] #Accuracy Score
     print("Accuracy Score: ",r2_score(y_test, pred)*100,"%")

     Accuracy Score:  88.50492246220446 %
```

Figure 4.5.2 accuracy of the model

The accuracy of the model is 88.50%.

```
from sklearn import metrics
from sklearn.metrics import r2_score

print("Mean Absolute Error (MAE): ",metrics.mean_absolute_error(y_test,pred))
print("Mean Squared Error (MSE): ",metrics.mean_squared_error(y_test,pred))
print("Root Mean Squared Error (RMSE): ",np.sqrt(metrics.mean_squared_error(y_test,pred)))

Mean Absolute Error (MAE):  145.44444444444446
Mean Squared Error (MSE):  32425.666666666668
Root Mean Squared Error (RMSE):  180.07128218199222
```

Figure 4.5.3 decision tree **modal evaluation result**

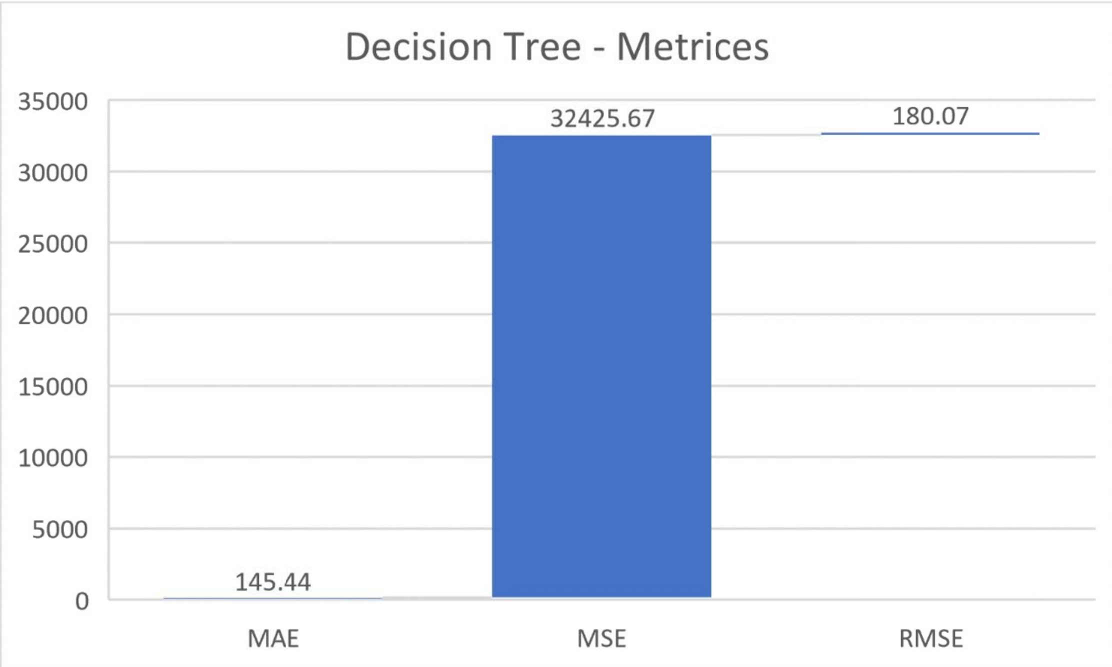


Figure 4.5.4 Decision Tree Regression Metrices

4.6 Confusion Matrix

“A performance evaluation's purpose is to assess the precision with which a classification algorithm system has been created. One technique for assessing performance is a confusion matrix.

Actual	**Prediction**	
	Economic	**Non-Economic**
Economic	TE(TrueEconomic)	TE(TrueEconomic)
Non Economic	FN(False Non-Economic)	TN(True Non-Economic)

Table 4.6.1 confusion matrix table

Where:

1. The quantity of economic data that the system expects to be true is known as TE (True Economic).
2. The quantity of non-economic data that the system expects to be genuine is known as TN (True Non-Economic).
3. The quantity of economic data that the system predicts to be erroneous is known as FE (False Economic).
4. The quantity of non-economic data that the system predicts to be false is known as FN (False Non-Economic)”[19].

1. *Precision*

The accuracy of a system in foretelling economic outcomes is measured by its precision. Equation (9) is used to get precision.

Precision = $/TE+FP$ (9)

2. Recall

A case of economic data that the system anticipates to be true is called recall. Equation (10) is used to calculate recall.

Recall = $TE\ /TE+FN$ (10)

3. *Accuracy*

Accuracy is determined by comparing accurately predicted data to all data in a single case. Equation (11), which calculates accuracy, says:

Accuracy = $TE+TN\ /TE+FN+TN+FE$ (11)

4. *ErrorRate*

The error rate is a comparison between all data in one case and expected incorrect data. Equation (12) is used to calculate it.

Error Rate = $FE+FN\ /TE+FN+TN+FE$ (12)

TABLE 4.6.2. Price Classification

Comparison	***Class***
Farmer's Selling Price > Production price	*Economic*
Farmer's Selling Price ≤ Production price	*Non-Economic*

Comparison table between the models

S.no	Model name	Accuracy
1.	**Linear Regression**	96.71%
2.	**Multiple Regression**	90.94%
3.	**KNN classifier**	39.28%
4.	**KNN Regression**	96.94%
5.	**Gradient Boosting Regression**	84.41%
6.	**Decision Tree**	88.50%

Table 4.6.3Comparison table of accuracy

Comparsion table of MAE, MSE and RMSE

Metrices	MAE	MSE	RMSE
Linear Regression	61.82	7324	85.58
Multi - Linear Regression	149.9	33559.14	183.58
KNN Regression	124.86	29211.43	170.91
Gradient Boosting	109.6	17307.49	131.57
Decision Tree	124.86	29211.43	170.91

Table 4.6.4Comparsion table of MAE, MSE and RMSE

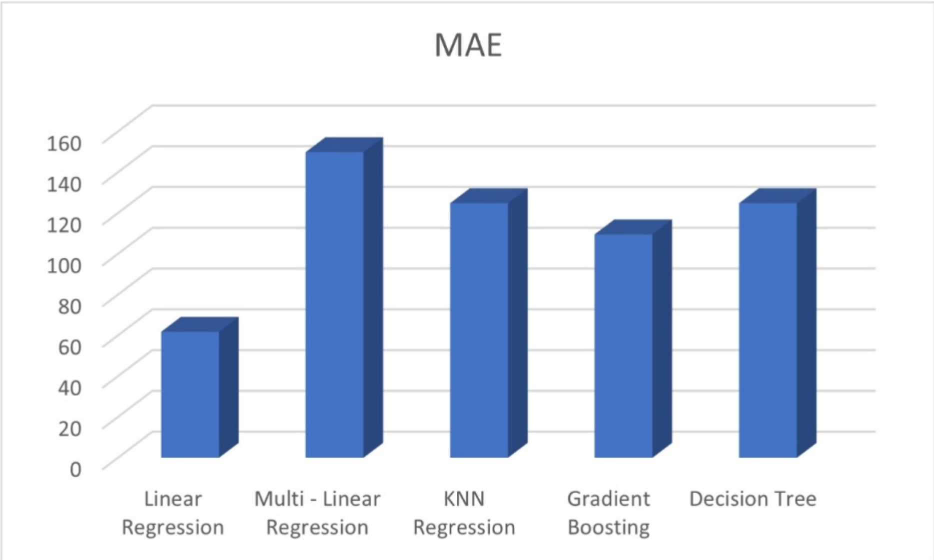


Figure 4.6.1 comparison of MAE between the models

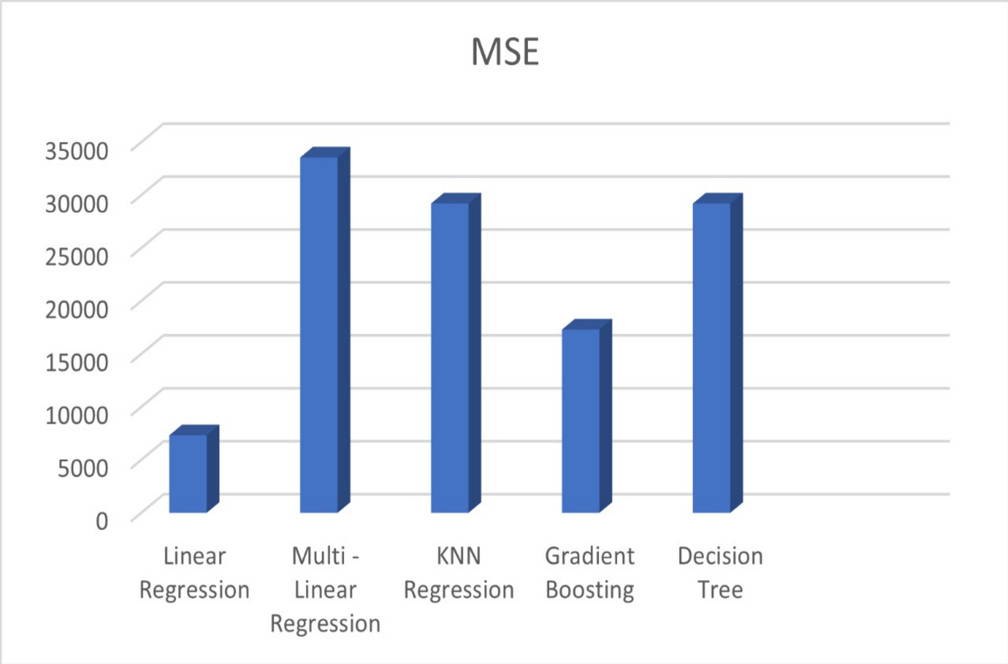


Figure 4.6.2 comparison of MSE between the models

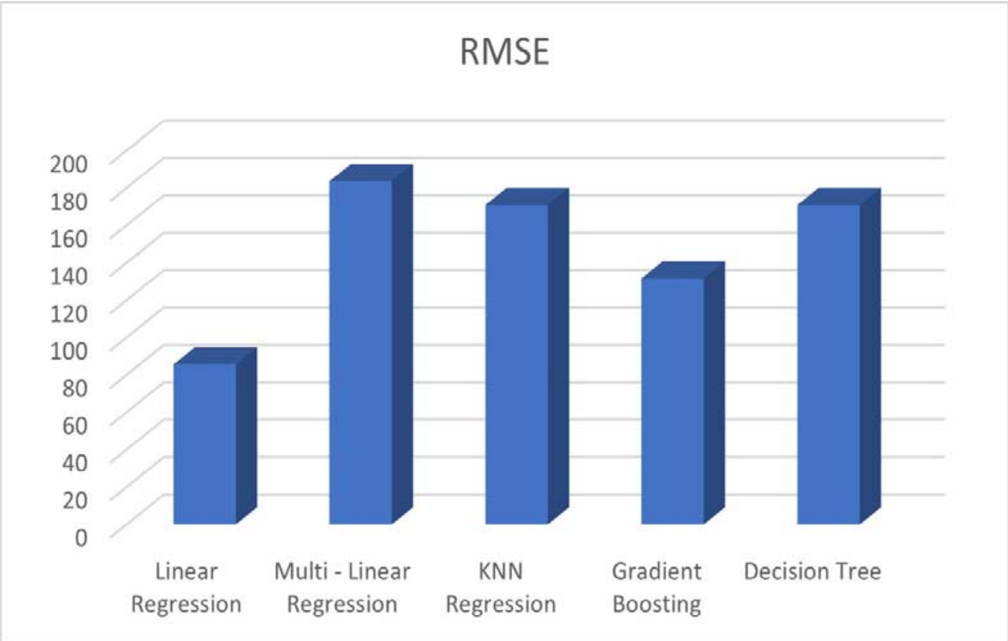


Figure 4.6.3 comparison of RMSE between the models

Chapter 5

Conclusion and Future Scope

Our programme will be improved in the future by adding new features. These papers just serve as decision-support tools for farmers. The Indian economy is operating in a free market, with market forces determining commodity prices based on supply and demand factors. Commercial crop prices are influenced by both domestic supply and demand dynamics as well as global supply and demand patterns. In this case, it is crucial to predict future prices (harvest month pricing) and communicate them to farmers so they may make well-informed decisions. This will influence the crop they choose. The paper projected crop prices using historical yearly prices.

Similar to any other method, this one does not ensure accurate forecasts. However, the model is useful and has been successfully applied to future forecasting. AlmemaychuAmera (2002), Punitha (2007), Darekar et al. (2016), and Jalikatti& Patil (2015) all employed a similar model to predict the pricing and arrivals of agricultural commodities and come to their own conclusions. The KNN Regression model's price predictions showed that both cotton prices and crop demand would rise in the coming years. This prediction is based on historical data and a model, so it's possible that the actual market price will differ from what was predicted.

Future scope

Future modifications to the model suggested in this paper could include a feature that forecasts crop rotations. The decision of which crop to grow will now also depend on which crop was harvested in the previous cycle, ensuring maximum output. In addition, the Crop Suitability Predictor model can take into account crop demand and supply as well as other economic variables like farm harvest prices and retail pricing. This would give a comprehensive forecast that took into account economic issues in addition to environmental and geographic ones.

References

1. Smith, Adam (1776) 1952 *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Chicago: Encyclopedia Britannica.

2. N. HemageethaandG.M.Nasira, “Radial Basis Function Model for Vegetable Price Prediction”, proceedings of the International Conference on Pattern Recognition, informatics and Mobile Engineering, 2013.

3. Yung-Hsing Peng, Chin- Shun Hsu, and Po-Chuang Huang, “An Investigation of Spacial approaches for Crop Price Forecasting in Different Taiwan Markets”, Conference on technologies and Application of Artificial Intelligence (TAAI),IEEE 2015.

4. Rohit Kumar Rajak, Ankit Pawar. MitaleePendke, Pooja Shinde. Suresh Rathod and AvinashDevare. "Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique," International Research Journal of Engineering and Technology (IRJET). 2017.

5. Chenhao Wang and QiangGao, “High and Low Prices Prediction of Soybean Futures with LSTM Neural Network,” 9th International Conference on Software Engineering and Service Science (ICSESS) IEEE ,2018.

6. Ashwini Darekar and A Amarender Reddy, “cotton price forecasting in major producing states,” Econ. Affair,2017.

7. ZeelDoshi ,SubhashNadkarni,Rashi Agrawal and Prof. Neepa Shah, “AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms,” Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

8. Lakshmi. N, Priya. M, Mrs. Sahana Shetty and Mr. Manjunath C.R , “Crop Recommendation System for Precision Agriculture,” International Journal for Research in Applied Science & Engineering Technology (IJRASET),2018.

9. Yongli Zhang , and Sanggyun Na, “A Novel Agricultural Commodity Price Forecasting Model Based on Fuzzy Information Granulation and MEA-SVM Model,” Hindawi Mathematical Problems in Engineering, 2018

10. A. Amarender Reddy, “Price Forecasting of Tomatoes,” International Journal of Vegetable Science,2018.

11. Tao Xiong, Chongguang Li and Yukun Bao, “Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China,” Neurocomputing Elsevier,2018.

12. Koki Yamamoto, Kazuhiro Seida, Junna Nishiyama , Kaito Hayashi , Suzuki Daishi and MikikoSode Tanaka, “Predicting System of Harvest Time And Yield of Tomato,” IEEE 7th Global Conference on Consumer Electronics (GCCE),2018.

13. Gerasimos Rigatos, PierluigiSiano, Taniya Ghosh and Yi Ding “Forecasting of commodities prices using a multi-factor PDE model and Kalman filtering,” IET Cyber-Physical Systems: Theory & Applications,2018.

14. Yuchen Weng , Xiujuan Wang , Jing Hua , Haoyu Wang, Mengzhen Kang and Fei-Yue Wang, “Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler,” IEEE Transactions on Computational Social Systems,2019.

15. Asma Hasifa Nurcahyono, Fhira Nhita, DeniSaepudin and AnnisaAditsania,“Price Prediction of Chili in Bandung Regency Using Support Vector Machine (SVM) Optimized with an Adaptive Neuro- Fuzzy Inference System (ANFIS),” 7th International Conference on Information and Communication Technology (ICoICT),2019.

16. Dabin Zhang, Shanying Chen , Liwen Ling, And Qiang Xia “Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons,” IEEE Access February 14, 2020

17. D.H.WolpertandW.G.Macready,‘‘Nofreelunchtheoremsforoptimiza-tion,’’ *IEEE Trans Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

18. https://www.kaashivinfotech.com

19. Submitted to National Institute of Technology

20. Ijsrst.com

21. www.scienpress.com

22. www.javatpoint.com

23. www.Coursehero.com

24. www.ijitee.org

25. www.investopedia.com

26. https://raw.githubusercontent.com/ParthGupta11/Wheat-demand-predictor/master/model_data.csv

Crop Price Predition

by Abhishek Dhull

Submission date: 29-Jul-2022 10:22AM (UTC+0530)
Submission ID: 1876446340
File name: project_-_Copy_1.docx (491.57K)
Word count: 8915
Character count: 47926

Crop Price Predition

ORIGINALITY REPORT

8%	3%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Asma Hasifa Nurcahyono, Fhira Nhita, Deni Saepudin, Annisa Aditsania. "Price Prediction of Chili in Bandung Regency Using Support Vector Machine (SVM) Optimized with an Adaptive Neuro-Fuzzy Inference System (ANFIS)", 2019 7th International Conference on Information and Communication Technology (ICoICT), 2019 Publication	1%
2	www.kaashivinfotech.com Internet Source	1%
3	Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Neepa Shah. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018 Publication	<1%
4	fr.slideshare.net Internet Source	<1%