Predicting Agricultural Produce Prices Using Deep Learning

Submitted in partial fulfillment of the requirements
of the degree of

BACHELOR OF ENGINEERING
In
COMPUTER ENGINEERING

By

Group No: 32

Roll No.	Name
1704009	Atharva Barwe
1704019	Milan Chandiramani
1704031	Chinmayi Dabholkar

Guide:

PROF. VAIBHAV AMBHIRE

(Assistant Professor, Department of Computer Engineering, TSEC)

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Computer Engineering Department
Thadomal Shahani Engineering College
University of Mumbai
2020-2021

CERTIFICATE

This is to certify that the project entitled **"Predicting Agricultural Produce Prices Using Deep Learning"** is a bonafide work of

Roll No.	Name
1704009	Atharva Barwe
1704019	Milan Chandiramani
1704031	Chinmayi Dabholkar

Submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"BACHELOR OF ENGINEERING"** in **"COMPUTER ENGINEERING"**.

Prof. Vaibhav Ambhire
Guide

Dr. Tanuja Sarode
Head of Department

Dr.G.T.Thampi
Principal

Project Report Approval for B.E

Project report entitled "**Predicting Agricultural Produce Prices Using Deep Learning**" by

Roll No.	Name
1704009	Atharva Barwe
1704019	Milan Chandiramani
1704031	Chinmayi Dabholkar

is approved for the degree of ***"BACHELOR OF ENGINEERING" in***

"COMPUTER ENGINEERING".

Examiners

1.--

2.--

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken when needed.

1) ______________________

Atharva Barwe , 1704009

(Name of student and Roll No.)

2) ______________________

Milan Chandiramani,1704019

(Name of student and Roll No.)

3) ______________________

Chinmayi Dabholkar,1704031

(Name of student and Roll No.)

Date: 26th November, 2020

Abstract

One of the most pressing issues in India is the large number of farmer suicides. More than 11,000 Indian farmers committed suicide in 2016, according to the National Crime Records Bureau. The Indian government has been taking measures to solve this problem for several years. Government markets buy crops at the minimum support prices but only upto a minimum quota. As a result, the farmers end up selling their crops to third-party vendors that don’t guarantee minimum support prices, and the farmers don’t make a profit. One of the main challenges of this project is that future produce prices have a long-term temporal dependence on past prices (e.g., the price of tomatoes in August 2019 may depend on their price in August 2018) and a spatial dependence on the prices at nearby markets (e.g., prices at nearby markets may be similar, as opposed to geographically distant markets), and thus, it is important to develop prediction models which can explicitly capture this spatio-temporal dependence. While previous work presents algorithms to predict future produce prices, they (i) do not explicitly consider the spatio-temporal dependence of future prices on past data; and as a result, (ii) they rely on classical ML prediction models (e.g., decision trees) which often perform poorly when applied to spatio-temporal datasets. These shortcomings limit the accuracy (and hence, usability) of these methods in this project. We propose a deep learning model that can estimate the price of a certain crop at a certain market in India. The price predicted by the model can help farmers to determine where and when they should sell their produce. The primary motivation behind this project is to maximize the profit of an individual farmer thus decreasing their likelihood of suffering a loss.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

India is a country where the major source of living and economy is through agriculture and agricultural industry. Agriculture is the main pillar of the economy in our country. Government is also facilitating farmers with many features such as subsidy in fertilizers, cancellation of their interest for loans etc. Most of the families rely on Agriculture. Country's Gross Development predominantly leans on Agriculture. 60% of the land is utilized for Agriculture to adequate the requirements of the Country's population. To meet the requirements, modernization in Agricultural practices is required. Prediction of prices in agricultural commodities has always been a major problem for the farmers. In the recent years there has been an inconsistency in the prices of multiple crops which in turn has increased the menace encountered by the farmers. The main purpose of the system is to predict effective crop prices.In the recent times, most of the farmers are deprived of the knowledge about the various breeds crops, season of sowing seeds, cultivation methodologies, cultivation cost and other conditions. The drop in the employment sector in the agricultural field is shown in the following figure.

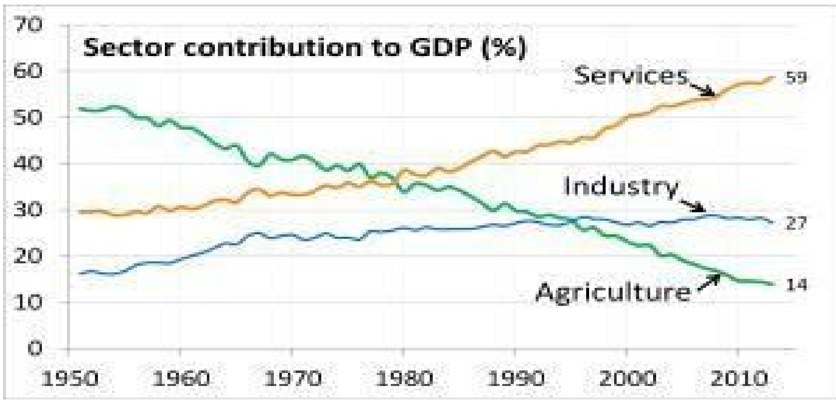

Figure 1.1 Graph depicting the drop in agricultural section contribution to GDP.

Farmers aren't able to get desired price for their crops and that's why the suicide rate is increasing with every passing year. Most farmers are driven to suicide due to an inability to sell their produce at desired profit levels, which is caused by the widespread uncertainty/fluctuation in produce prices resulting from varying market conditions. Also due these factors, the crop loss has increased and hence the famers suicide has increased as shown in the figure.

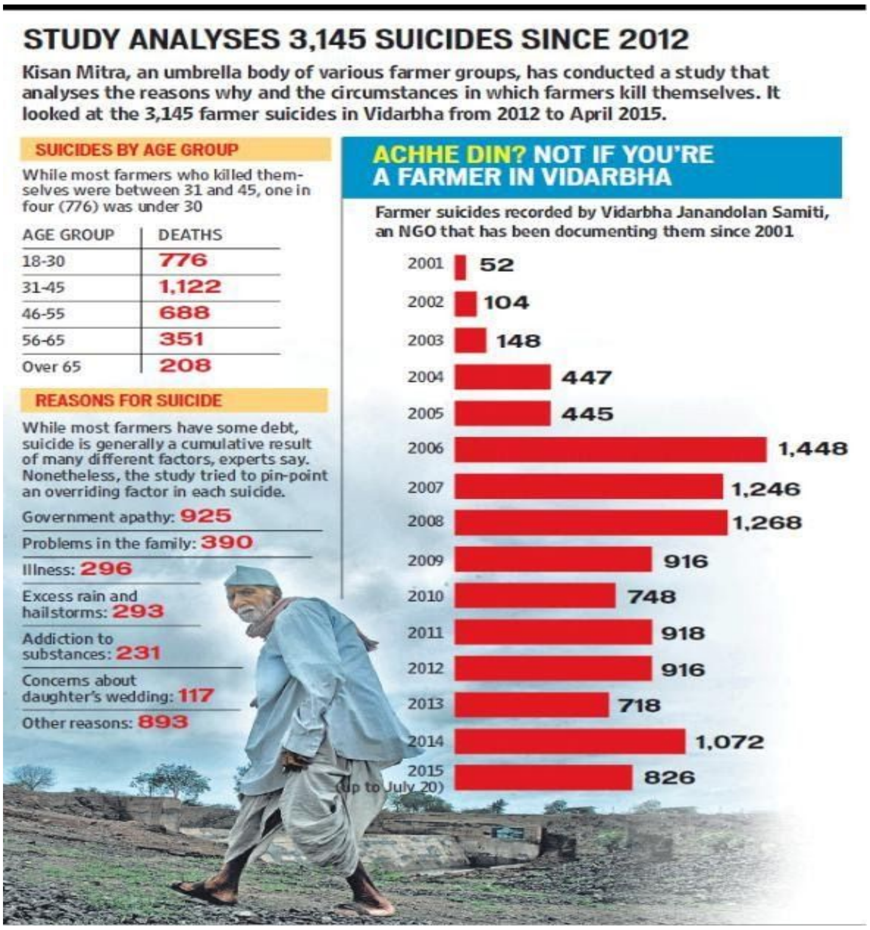

Figure 1.2 Increase In Farmer's suicide rates (Courtesy: Hindustan Times)

1.2 Aims & Objectives

The principal aim of this 4th year group project is to design and develop a model which will help to maximise the farmer's profit by predicting effective crop pices. The aim is also to propose a new framework and develop a system to make some advances towards more efficient price prediction.The objective of this project is to build a system which provides efficient and effective price prediction features. To achieve our objective we are proposing a model similar to PECAD (Price Estimation for Crops using the Application of Deep Learning), a novel neural network architecture to predict future prices of agricultural produce. First, it collects real-world prices and (produced) volume of different crops at approximately 1,350 agricultural markets in India over a period of 11 years from Agmarknet.gov.in1 (an official Indian government administered website). Second, PECAD preprocesses this raw dataset via state-of-the-art imputation (and other) techniques to account for missing data entries. Third, using this data as input, PECAD proposes a novel neural network architecture inspired by the wide and deep learning paradigm.

1.3 Scope

1.3.1 Scope and Future Work

The agricultural commodity prices have a volatile nature which may increase or decrease inconsistently causing an adverse effect on the economy. The work carried out here for predicting prices of agricultural commodities is useful for the farmers because of which they can sow appropriate crops depending on its future price. Agriculture products have seasonal rates; these rates are spread over the entire year. If these rates are known/alerted to the farmers in advance, then it will be promising on ROI (Return on Investments).

The scope of the project is to suggest the best crops based on the circumstances of the environment so that the farmers can bid accordingly for the prices and thus gain maximum profit. Price estimation in agriculture is mandatory to forecast the market price for the chosen commodities, and it is supportive for farmers to schedule their crop cultivation operations and thus they gain more profit. Market price prediction is mandatory for both private and public sectors to plan and execute agricultural development programs to steady the commodities market price. Since this is the present scenario, many of the farmers have insufficient understanding of the fresh plants and are not fully conscious of the advantages they gain from farming them. In addition, by knowing and predicting crop efficiency in a multitude of environmental circumstances, farm productivity can be improved. As one who should understand how much he would expect for their plants, predicting prices is a very significant problem for many farmers. Over the previous few years, price prediction has been produced by assessing farmers' experience on a specific crop and field. Suppose they have prior year information accessible in which distinct respective price projections are recorded and these recorded price projections are used to classify future price projections . By using these predictive models, the government can do agricultural development planning to stabilize the respective product prices.

The future work of this model involves building this model in the regional languages, so that it would be more comfortable for farmers as many of them are illiterates. Further it can be improved by adding certain add-ons such as details about crop diseases, information about micronutrients etc. This helps to reduce the crop loss by properly educating the farmers with the crops details and their requirements.

1.3.2 Challenges

Challenges are the major basis which imminent the negative impacts on current project. Some of the challenges faced during crop yield prediction are:

1. Choosing appropriate dataset, after choosing dataset tuning of the parameters which makes the project more efficient to get the desired results.
2. Model must be trained by taking consideration of less computational efficiency and power.
3. Increase in error rate due to dynamically changing the environment.
4. Problem is spatio-temporal based implying better as well as experimental analysis of data and choosing the right algorithm/model to achieve as better accuracy as possible.

Chapter 2

Literature Survey

2.1 Survey Existing Systems

2.1.1 Decision Making Support System for Prediction of Prices in Agricultural Commodity

India is an agriculture based country and farmer community is the backbone of the agriculture sector. Agribusiness is one of the important segments of the agriculture sector. The model aims to minimize agribusiness risk by deploying the agribusiness knowledge to the farmer community. The model is intended at predicting agriculture commodity market price using deep learning. Such generated knowledge helps the farmer community in their decision making and thereby achieving their ultimate goal of profit making. The prices of agricultural commodities have an unstable nature which may rise or fall differently causing negative effects on the economy. The work completed here for predicting costs of horticultural wares is helpful for the farmers as they can sow crops depending upon its future cost. Farming items have regular rates; these rates are spread over the whole year. In the event that these rates are

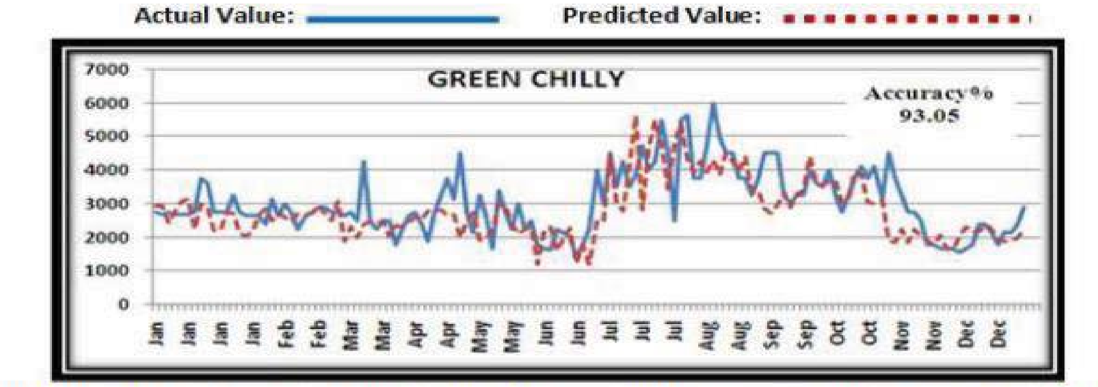

Figure 2.1 Market Price Prediction for Agriculture Commodity Green Chilly

known to farmers ahead of time, it will be guaranting on Rate on Investments (ROI). Horticultural specialists can pursue these charts and anticipate advertise rates which can be informed to farmers. The outcomes will be given dependent on the area of the clients of this application [1].

2.1.2 Data Mining in Agriculture on crop price prediction:Techniques and Applications

According to Kaur Manpreet (2014) – “The price change of Crude Oil will bring a change in prices of Agricultural commodities.”[5]

The Government of India has set a Minimum Support Price (MSP) for some crops like cereals and pulses that help the farmers to get at least minimum price for their crops.

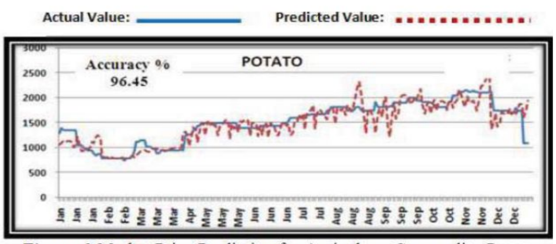

Figure 2.2 Market Price Prediction for Agriculture Commodity Potato

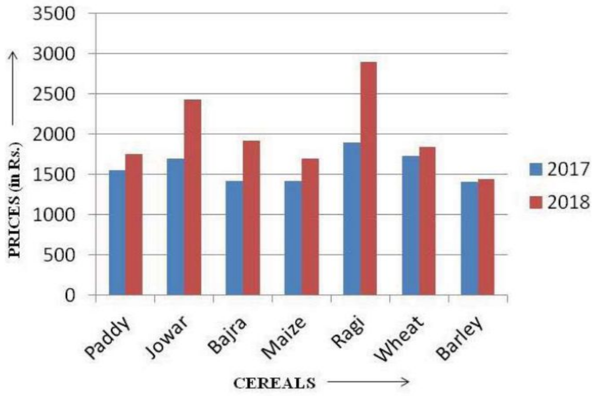

Figure 2.3 Comparison between cereals prices of year 2017 & 2018.

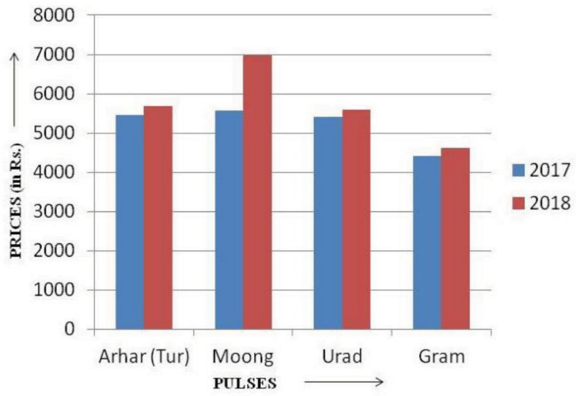

Figure 2.4 Comparison between pulses prices of year 2017 & 2018.

This comparison between the prices can help the farmers to predict in a better way and bid accordingly.

Commodity	Prices in Rs. (per quintal)	
	2017	2018
Cereals		
Paddy	1550	1750
Jowar	1700	2430
Bajra	1425	1925
Maize	1425	1700
Ragi	1990	2897
Wheat	1735	1840
Barley	1410	1440
Pulses		
Arhar (Tur)	5450	5675
Moong	5575	6975
Urad	5400	5600
Gram	4400	4620

Table 1 – Prices of Cereals and Pulses for year 2017 & 2018

Table 2.1 Prices of cereals and pulses for year 2017 & 2018.

2.1.3 A new method of large-scale short-term forecasting of agricultural commodity prices

According to Huaili Wu (2017) – “ARIMA model, neural networks and PLS regression can be used for price forecasting for weekly basis.”[6]

2.1.4 Artificial Neural Network Methodology for Modelling and Forecasting Maize Crop Yield

According to Prajenshu (2008) – “Multilayered feedforward ANN a type of Artificial Neural Network technique can be used Maize crop yield forecasting.”[7]

2.1.5 A survey on data mining techniques in agriculture

According to M.C.S Geetha (2007) – “Information Technology can play a major role in crop yield prediction and help in better decision making.”[8]

2.1.6 A brief survey of data mining techniques applied to agricultural data

According to Hetal Patel (2014) – “Data mining, data mining techniques and data mining Applications plays a major role in crop prediction.”[9]

2.1.7 Analysis of Data Mining Techniques for Agriculture Data

According to Djodiltachoumy (2016) – “Neural networks, bi-clustring, k-means, k-nearest neighbour Naive Bayes Classifier and support vector machine are useful techniques for prediction in agricultural commodities”[10]

2.2 Limitation Existing Systems or Research Gaps

A few implementation challenges need to be solved when the model gets deployed by non-profit agencies working with indebted farmers. First, a model's predictive performance can potentially be improved by incorporating historical weather patterns, which can play a role in determining future crop supply (and hence, the future crop price). However, deep learning methods are rarely used to model weather in the real-world, as physical models are far more accurate at predicting future weather. Thus, the model needs to be integrated with physical weather prediction models (as part of future work). Further, sophisticated deep learning approaches to predicting future produce prices may raise suspicions among low-literate farmers. Public awareness campaigns in the agencies working with this program would help overcome such fears and to encourage participation. Also, non-profit agencies often do not prioritize spending their limited resources to buy sophisticated computer hardware (to train and run the model). Thus, we propose deploying the model as a stand-alone web service that the agencies could use without our intervention. Finally, model represents a single piece of the puzzle that needs to be solved for preventing farmer suicides, there are many other pieces. For example, a model's successful deployment depends crucially on availability of long-term crop pricing and volume patterns. While Agmarknet.gov.in makes this information available for Indian markets, there are no analogous data repositories for other developing countries.

2.3 Problem Statement

To design a model to predict market prices for agricultural produce. The determining factors are location, type, variety and quality of crop, volume patterns, historical pricing, etc. The input will be processed by a deep neural network. Performance of the model will be maximised by tuning the hyperparameters. The output of the model will be the predicted price for a particular crop in a particular market. The dataset will be web-scraped from the Indian government website ‘Agriculture Marketing’.

Chapter 3

Proposed System

3.1 Methodology

Introduction

There are different types of research methods either in social science, management, medical, engineering among others. In all the fields of study, there are various research methods available and understanding of these methods will assist an individual to choose the right research methodology in his or her research exercise. There are many ways to categorize different types of research. The words used to describe the research depends on the discipline and field. Generally, the form ones research (types of research methods) approach takes will be shaped by the followings:

- The type of knowledge one aims to produce
- The type of data one will collect and analyze
- The sampling methods, timescale and location of the research.
- Types of research methods

For our project, we have considered experimental research.

Experimental Research

Experimental research is one of the research methods, and it is any research conducted with a scientific approach, where a set of variables are kept constant while the other set of variables are being measured as the subject of experiment. There are times when you don't have enough data to support your decisions. In such situations, you need to carry out experiments to discover the facts. Experimental research can gather a lot of data that can help one make better decisions or help discover unknown knowledge.

Steps for Research

The methodology involves:

- Review of Literature
- Data Collection and Preprocessing (as explained in Chapter 4, Subsections 4.1.1 and 4.1.2)
- Algorithm Selection and Model Creation
- Model Training and Testing
- Model Improvisation
- Reiteration of last 3 steps until acceptable benchmark is reached

3.2 Algorithm

Artificial Neural Networks are the computational models that are inspired by the human brain. Many of the recent advancements have been made in the field of Artificial Intelligence, including Voice Recognition, Image Recognition, Robotics using Artificial Neural Networks. Artificial Neural Networks are the biologically inspired simulations performed on the computer to perform certain specific tasks like:

1. Clustering

2. Classification

3. Pattern Recognition

Artificial Neural Networks, in general is a biologically inspired network of artificial neurons configured to perform specific tasks.These biological methods of computing are considered to be the next major advancement in the Computing Industry.

3.2.1 Design Details

Basic Structure of ANNs :

The idea of ANNs is based on the belief that working of human brain by making the right connections, can be imitated using silicon and wires as living neurons and dendrites.

The human brain is composed of 86 billion nerve cells called neurons. They are connected to other thousand cells by **Axons.** Stimuli from the external environment or inputs from sensory organs are accepted by dendrites. These inputs create electric impulses, which quickly travel through the neural network. A neuron can then send the message to another neuron to handle the issue or does not send it forward.

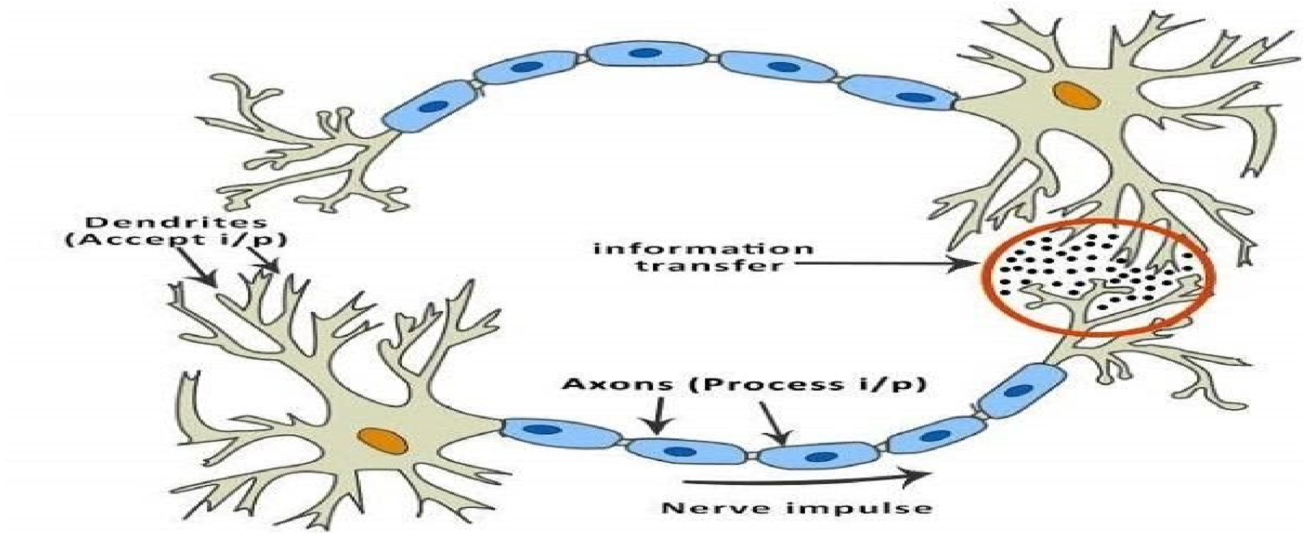

Fig 3.1 Basic Structure Of Artificial Neural Network

ANNs are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value.

Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values.

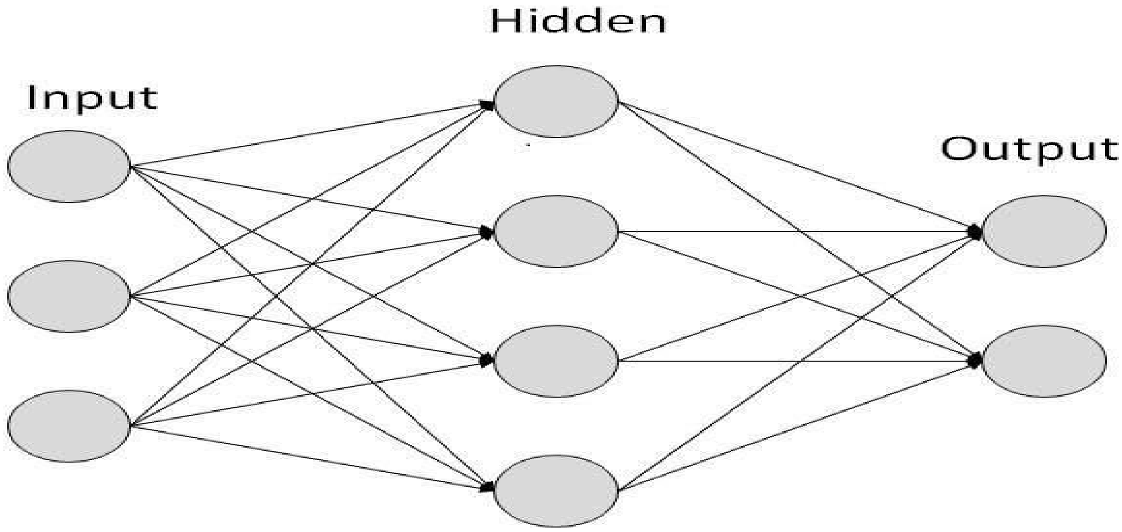

Fig 3.2 Illustration Of Simple Artificial Neural Network

Types of Artificial Neural Networks:

There are two Artificial Neural Network topologies : FeedForward and Feedback.

Feedforward ANN

In this ANN, the information flow is unidirectional. A unit sends information to another unit from which it does not receive any information. There are no feedback loops. They are used in pattern generation/recognition/classification. They have fixed inputs and outputs.

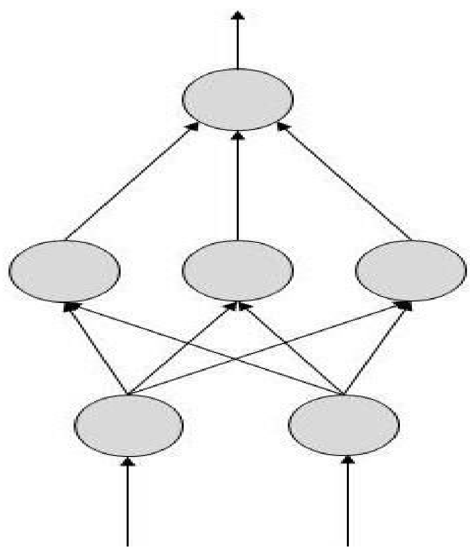

Fig 3.3 Feedforward Artificial Neural Network FeedBack ANN

Feedback ANN

Here, feedback loops are allowed. They are used in content addressable memories.

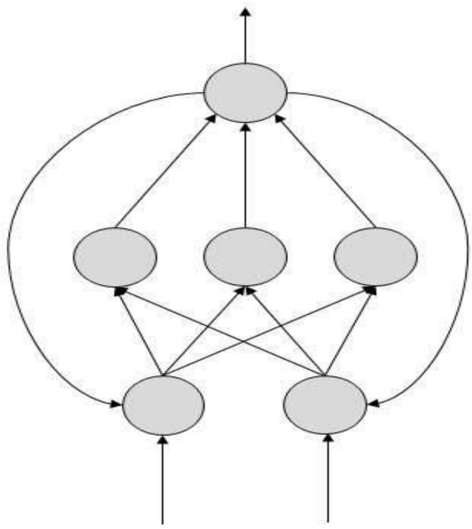

Fig 3.4 Feedback Artificial Neural Network

Working of ANNs :

In the topology diagrams shown, each arrow represents a connection between two neurons and indicates the pathway for the flow of information. Each connection has a weight, an integer number that controls the signal between the two neurons.

If the network generates a “good or desired” output, there is no need to adjust the weights. However, if the network generates a “poor or undesired” output or an error, then the system alters the weights in order to improve subsequent results.

ANN Design for Price Prediction:

The flowchart in Figure 3.5 shows the ANN layers for the model used in our project. As it is visible, the model consists of 6 hidden layers and 1 output layer. The inputs are given to the model which are passed to the first hidden layer. The first hidden layer consists of 128 neurons which are densely connected. The output of this layer is given to the second hidden layer which consists of 64 densely connected neurons. The output of the second hidden layer is given to the third hidden layer which consists of 32 densely connected neurons. Then output of the third hidden layer is transferred to the fourth hidden layer which consists of 16 densely connected neurons. The neurons after learning transfer their outputs to the fifth hidden layer consisting of 8 densely connected neurons. The output of this hidden layer is given to the output layer consisting of 3 densely connected neurons. The activation function for all hidden layers is a rectified linear unit(ReLU) function.

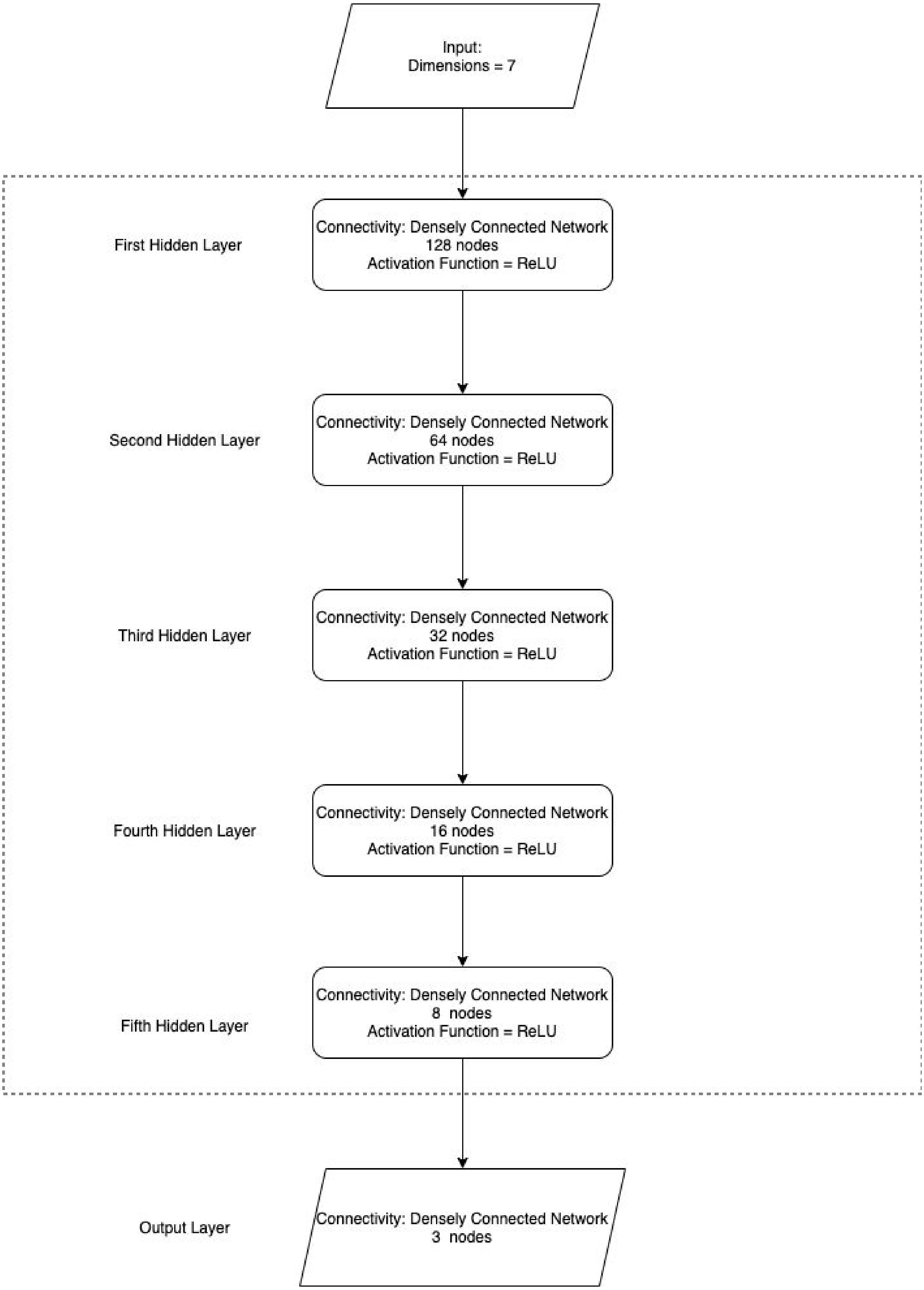

Figure 3.5 ANN Design For Price Prediction

3.2.2 Input Output Design

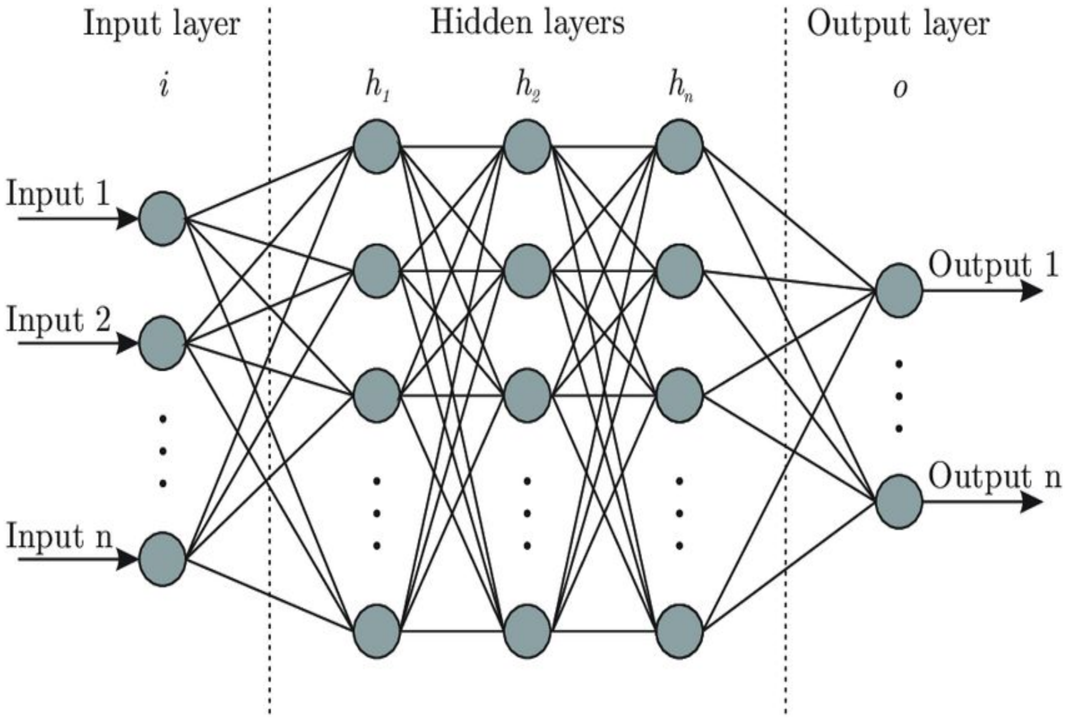

Figure 3.6 Input Output ANN Design Structure

Input neurons:

This is the number of features your neural network uses to make its predictions.

The input vector needs one input neuron per feature. For tabular data, this is the number of relevant features in your dataset. You want to carefully select these features and remove any that may contain patterns that won't generalize beyond the training set (and cause overfitting).

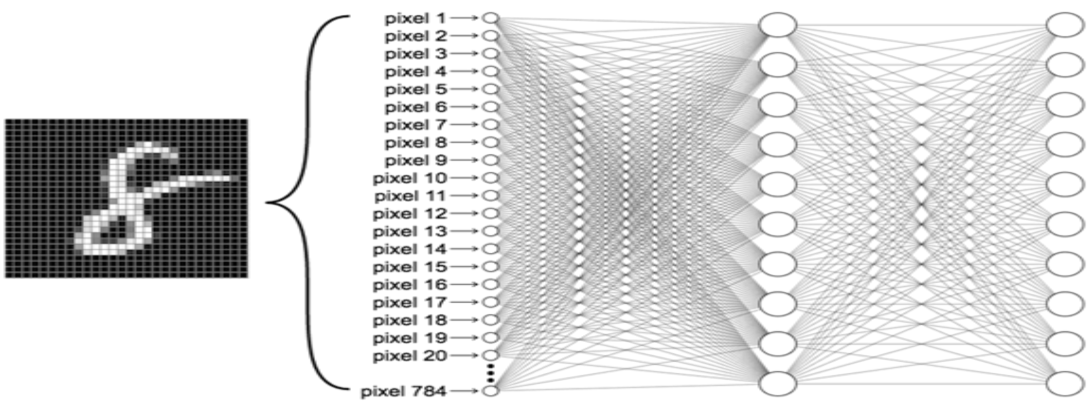

Figure 3.7 Input Neuron

Output neurons:

This is the number of predictions you want to make.

Regression: For regression tasks, this can be one value (e.g. housing price). For multivariate regression, it is one neuron per predicted value (e.g. for bounding boxes it can be 4 neurons — one each for bounding box height, width, x-coordinate, y-coordinate).

Classification: For binary classification (spam-not spam), we use one output neuron per positive class, wherein the output represents the probability of the positive class. For multi-class classification (e.g. in object detection where an instance can be classified as a car, a dog, a house etc.), we have one output neuron per class, and use the softmax activation function on the output layer to ensure the final probabilities sum to 1.

Since the given problem is a regression problem, we have one neuron per output, i.e., the output layer consists of 3 neurons.

Hidden Layers and Neurons per Hidden Layers :

The number of hidden layers is highly dependent on the problem and the architecture of your neural network.

In general, using the same number of neurons for all hidden layers will suffice. For some datasets, having a large first layer and following it up with smaller layers will lead to better performance as the first layer can learn a lot of lower-level features that can feed into a few higher order features in the subsequent layers.

For the problem statement under study,
the first hidden layers have 128 neurons,
followed by the next hidden layer having 64 neurons,
followed by the next hidden layer having 32 neurons,
followed by the next hidden layer having 16 neurons,
followed by the next hidden layer having 8 neurons,
followed by the output layer having 3 neurons to produce minimum price, maximum price and modal price.

Chapter 4

Implementation

4.1 Dataset Description

Agricultural Marketing Information Network (AGMARKNET) was launched in March 2000 by the Union Ministry of Agriculture. The Directorate of Marketing and Inspection (DMI), under the Ministry, links around 7,000 agricultural wholesale markets in India with the State Agricultural Marketing Boards and Directorates for effective information exchange. This e-governance portal agmarknet, implemented by National Informatics Centre (NIC), facilitates generation and transmission of prices, commodity arrival information from agricultural produce markets, and web-based dissemination to producers, consumers, traders, and policy makers transparently and quickly.

The agmarknet website (http://www.agmarknet.nic.in) is a G2C e-governance portal that caters to the needs of various stakeholders such as farmers, industry, policy makers and academic institutions by providing agricultural marketing related information from a single window. The portal has helped to reach farmers who do not have sufficient resources to get adequate market information. It facilitates web- based information flow, of the daily arrivals and prices of commodities in the agricultural produce markets spread across the country. The data transmitted from all the markets is available on the agmarknet portal in 8 regional languages and English. It displays Commodity-wise, Variety-wise daily prices and arrivals information from all wholesale markets. Various types of reports can be viewed including trend reports for prices and arrivals for important commodities. Currently, about 1,800 markets are connected and work is in progress for another 700 markets. The agmarknet portal now has a database of about 300 commodities and 2,000 varieties.

4.1.1 Data Collection :.

We collected all our raw data on agricultural crops (produce) from Agmarknet.gov.in1, a website run by the Indian government's Ministry of Agriculture and Farmers Welfare, which contains daily price and volume data at 1352 agricultural markets across India for over twelve Agmarknet Portal (http://agmarknet.gov.in) which disseminates daily market information of various commodities.To create open equitable markets, price information should be shared with farmers. By providing open data on markets and market prices, farmers will be better positioned to negotiate at the farm gate, select crops, and select a distribution channel. To break the power of the middle-man, the federal government of India aims to facilitate trade at a federal level as opposed to locally or regionally. Traditionally trade was restricted to a limited number of markets, resulting in a lack of competition, high market fees and a long chain of intermediation leading to low prices for the farmer and high prices for the consumer. New laws have been introduced first at federal level, later at regional level, allowing farmers to trade at any market they like in India and not restricting them to the local market. This new policy is supported by an open data portal called AGMARKNET. This provides information on the daily prices and arrivals of over 300 commodities and 2000 varieties at 3245 markets throughout the country. It also provides information about grading, standardization, packaging and quality certification regulations of the different markets. The price data from the agmarknet data portal is provided as open data at the Indian Open Data Government Data portal.The data refers to State-wise, market-wise, variety-wise prices of various crops. It has the daily wholesale maximum price, minimum price and modal price.

4.1.2 Data Preprocessing :.

The data from agmarknet.gov.in had several features, some of which were not useful to our application. Columns like ‘District Name’ and ‘Grade’ were deleted from the dataframe since it served no further purpose after giving the latitudes and longitudes. Another simple modification was to extract the months and years, convert them to numbers from strings, and add columns for them.

The next step was to use one hot encoding to enable the neural net to make sense of the type of crop such as rice, bananas, and wheat. Since neural networks work with numbers, one hot encoding is applied to convert said strings to numbers.

```
In [8]:  1 data
Out[8]:
```

	Market Name	Commodity	Min Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal Price (Rs./Quintal)	Price Date	Latitude	Longitude
0	Aheri	Rice	4000.0	4000.0	4000.0	20-Feb-18	26.650204	82.008342
1	Aheri	Rice	1800.0	1800.0	1800.0	27-Dec-17	26.650204	82.008342
2	Aheri	Rice	3500.0	3500.0	3500.0	19-Aug-17	26.650204	82.008342
3	Aheri	Rice	2000.0	2000.0	2000.0	29-Jun-17	26.650204	82.008342
4	Aheri	Rice	2000.0	2000.0	2000.0	20-Jun-17	26.650204	82.008342
...	...	...	...	...	...	...	...	...
185032	Manora	Wheat	1500.0	1750.0	1600.0	22-Feb-16	16.514398	93.643163
185033	Manora	Wheat	1450.0	1550.0	1500.0	08-Jan-16	16.514398	93.643163
185034	Manora	Wheat	1550.0	1650.0	1600.0	04-Nov-15	16.514398	93.643163
185035	Manora	Wheat	1500.0	1650.0	1550.0	05-Oct-15	16.514398	93.643163
185036	Manora	Wheat	1550.0	1650.0	1600.0	01-Oct-15	16.514398	93.643163

185037 rows × 8 columns

Figure 4.1 Data Preprocessing

```
In [14]: one_hot=pd.get_dummies(data.Commodity)

In [15]: data=pd.concat([one_hot,data],axis=1)

In [16]: del data['Commodity']

In [17]: months={'Jan':1,'Feb':2,'Mar':3,'Apr':4,'May':5,'Jun':6,'Jul':7,'Aug':8,'Sep':9,'Oct':10,'Nov':11,'Dec':12}

In [18]: data['Month']=[months[month] for month in data['Month']]

In [19]: data=data.rename(columns={"Min Price (Rs./Quintal)":"Min Price","Max Price (Rs./Quintal)":"Max Price","Modal Price

In [20]: data
Out[20]:
```

	Banana	Rice	Wheat	Min Price	Max Price	Modal Price	Latitude	Longitude	Month	Year
0	0	1	0	4000.0	4000.0	4000.0	26.650204	82.008342	2	18
1	0	1	0	1800.0	1800.0	1800.0	26.650204	82.008342	12	17
2	0	1	0	3500.0	3500.0	3500.0	26.650204	82.008342	8	17
3	0	1	0	2000.0	2000.0	2000.0	26.650204	82.008342	6	17
4	0	1	0	2000.0	2000.0	2000.0	26.650204	82.008342	6	17
...	...	...	...	...	...	...	...	...	...	...
185032	0	0	1	1500.0	1750.0	1600.0	16.514398	93.643163	2	16
185033	0	0	1	1450.0	1550.0	1500.0	16.514398	93.643163	1	16
185034	0	0	1	1550.0	1650.0	1600.0	16.514398	93.643163	11	15
185035	0	0	1	1500.0	1650.0	1550.0	16.514398	93.643163	10	15

Figure 4.2 One Hot Encoding

Before feeding data into the model, it must be normalized. We used Standard Scaler for this purpose. The idea behind StandardScaler is that it will transform your data such that its distribution will have a mean value 0 and standard deviation of 1. In case of multivariate data, this is done feature-wise (in other words independently for each column of the data). Given the distribution of the data, each value in the dataset will have the mean value subtracted, and then divided by the standard deviation of the whole dataset (or feature in the multivariate case).

```
In [112]: from sklearn.preprocessing import StandardScaler
          from sklearn.model_selection import train_test_split
          from tensorflow.keras.models import Sequential
          from tensorflow.keras.layers import Dense, Activation

In [113]: X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.1,random_state=7)

In [114]: sc=StandardScaler()
          X_train=sc.fit_transform(X_train)

In [115]: X_test=sc.transform(X_test)
```

Figure 4.3 Standard Scaler

4.2 Preliminary Results

Working with a sample dataset of 1,85,000 entries, we were able to get a satisfactory result from the neural network. Since regression problems do not have an accuracy measure and instead, we instead used the metric 'Mean Squared Error'. The model was trained for 150 epochs beyond which there was no appreciable improvement in the MSE for the additional cost. As a method of roughly estimating how well the model was predicting the output, correlation was used to compare the minimum, maximum, and modal prices with those predicted by the net.

```
In [120]:
plt.plot(model.history.history['loss'])
plt.plot(model.history.history['val_loss'])
plt.ylabel('Mean Squared Error')
plt.xlabel('Epoch')
plt.legend(['Train','Test'],loc='upper left')
plt.show()
```

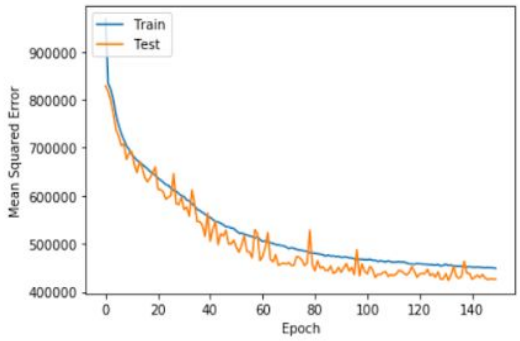

Figure 4.4 Loss Function (MSE)

```
In [138]:  Y_pred
Out[138]: array([[2407.9834, 5496.191 , 3562.5051],
                 [3622.401 , 4777.018 , 4050.0176],
                 [1787.1918, 3752.9622, 2853.5637],
                 ...,
                 [1088.9707, 1544.2379, 1367.9702],
                 [1744.5085, 2134.6057, 1933.1288],
                 [2800.389 , 3415.4124, 3090.8474]], dtype=float32)

In [139]:  Y_test
Out[139]: array([[2200., 5790., 3600.],
                 [3500., 4500., 4000.],
                 [1800., 3800., 2800.],
                 ...,
                 [1000., 1500., 1200.],
                 [3050., 3300., 3050.],
                 [2400., 2750., 2500.]])

In [133]:  pd.Series(Y_test.T[0]).corr(pd.Series(Y_pred.T[0]))
Out[133]: 0.8372635823977926

In [135]:  pd.Series(Y_test.T[1]).corr(pd.Series(Y_pred.T[1]))
Out[135]: 0.8052756239929528

In [134]:  pd.Series(Y_test.T[2]).corr(pd.Series(Y_pred.T[2]))
Out[134]: 0.8290058477080801
```

Figure 4.5 Correlation with True Values

Chapter 5

Summary

This system is proposed to provide help to the farmers for expecting the better amount for their crops and for predicting the best price for the crops.This also helps farmers to check previous prices of different commodities.This system will allow farmers to make better decisions for bidding the better prices for their crops in the market.

5.1 Future Prospects and Improvisations

The future prospects of this project are:

- Building a product based on this project which will make it available in the regional languages, thus, making it more comfortable for farmers as many of them prefer their local dialect. The product can have certain add-ons such as details about crop diseases, information about micronutrients etc. This helps to reduce the crop loss by properly educating the farmers with the crops details and their requirements.

- The dataset has a property of spatio temporal dependence , i.e. , it is dependent on space and time. As a result, ANN is unable to capture the pattern to its fullest. Implementation of deep wide neural networks of CNN and LSTM-RNN model can be a great choice to implement the problem statement at a broader scale.

- We believe that the model's predictive performance can potentially be improved by incorporating historical weather patterns, which can play a role in determining future crop supply and hence, the future crop price. However, deep learning methods are rarely used to model weather in the real-world, as physical models are far more accurate at predicting future weather. As a result, the model needs to have a data pipeline fetching weather based data in realtime. Thus, the model needs to be integrated with physical weather prediction models.

References

[1] Aman Vohra, Nitin Pandey, S.K. Khatri, “Decision Making Support System for Prediction of Prices in Agricultural Commodity,” in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 2019, pp.04-08

[2] Hangzhi Guo, Alexander Woodruff, Amulya Yadav, “Improving Lives of Indebted Farmers Using Deep Learning: Predicting Agricultural Produce Prices Using Convolutional Neural Networks”; 2019,pp.13-16

[3] Rachana P, Rashmi G,Shravani D,Shruthi N, Seema Kousar,”CROP PRICE FORECASTING SYSTEM USING SUPERVISED MACHINE LEARNING ALGORITHMS”,2019,pp.10-12

[4] Girish K.Jha ,Kanchan Sinha ,”Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System” in *2013 Indian Agricultural Research Institute,New Delhi,* 2013,pp. 05-09

[5] Kaur, Manpreet, Heena Gulati,Harish Kundra,”Data mining in Agriculture on crop price prediction: Techniques and Applications." in *2014 International Journal of Computer Applications* , 2014,pp. 02-07

[6] Wu, H., Wu, H., Zhu, M., Chen, W., & Chen, W,” A new method of large-scale short-term forecasting of agricultural commodity prices”,in *2017 Case of agricultural markets in Beijing. Journal of Big Data,*2017,pp. 03-09

[7] Krishna Singh, R.,” Artificial Neural Network Methodology for Modelling and Forecasting Maize Crop Yield”,2008,pp.06-07

[8] Geetha, M.,” A survey on data mining techniques in Agriculture” in *2015 International journal of innovative research in computer and communication engineering*,2015,pp 14-15

[9] Patel, H., & Patel, D,”A brief survey of data mining techniques applied to agricultural data” in *2014 International Journal of Computer Applications*, 2014,pp.16-27

[10] Djodiltachoumy, S.”Analysis of Data Mining Techniques for Agriculture Data.” in *2016 International Journal of Computer Science and Engineering Communications,* 2016,pp.32-41.

Acknowledgement

Foremost, we would like to express our sincere gratitude to our advisor and projectguide Prof. Vaibhav Ambhire for the continuous support in the study and research of this project, for his patience, motivation, enthusiasm, and immense knowledge. His dynamism, vision, sincerity and motivation have deeply inspired us. He has taught us the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under his guidance. We could not have imagined having a better guide and mentor for our BE project. We are really thankful to him.

We would also like to thank our Head of Department(HOD) Ms. Tanuja Sarode, for their encouragement, insightful comments, and hard questions. Our sincere thanks also goes to Mr. Anil Chhangiani, Ms. Darakshan Khan, Ms. Nabanita Mandal, Ms. Sachi Natu and Dr. Mayank Daswani, for offering the much required insights in our data collection and domain understanding. We would like to express our special thanks of gratitude to our principal who gave us the golden opportunity to do this wonderful project on the topic "Predicting Agricultural Produce Prices Using Deep Learning" , which also helped us in doing a lot of Research. We are overwhelmed in all humbleness and gratefulness to acknowledge our depth to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete.

Any attempt at any level can't be satisfactorily completed without the support and guidance of parents, friends and closed ones that support you with their prayers, empathy, and a great sense of humor. Finally, our thanks go to all the people who have supported us to complete the research work directly or indirectly.

Atharva Barwe

Milan Chandiramani

Chinmayi Dabholkar