Project Report
On

"Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks"

For Partial Fulfillment of B E (Computer Science & Engineering)
During Academic Year 2017-2018
Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal

By
Tejas Verghese [0111CS141111]
Technocrats Institute of Technology, Bhopal

Under guidance of
Asst. Prof. Mayank Pathak
Department of Computer Science & Engineering
Technocrats Institute of Technology (TIT)
Bhopal-462021

Session: Jan – May 2018

Technocrats Institute Of Technology

Bhopal (M.P.)

STATEMENT BY THE CANDIDATE

I state that work embodied in this Project Dissertation titled **"Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks"** forms my own contribution to the work carried out under the Guidance of Asst. Prof. Mayank Pathak at the Technocrats Institute of Technology (TIT), Bhopal - 462021. This work has not been submitted for any other Degree or Diploma of any University / Institute. Wherever references have been made to previous works of others, they have been appropriately acknowledged.

Tejas Verghese

Fourth Year, B.E.

Department of Computer Science & Engineering

Technocrats Institute Of Technology

Bhopal (M.P.)

CERTIFICATE

This is to certify that **Mr. Tejas Verghese [0111CS141111]**, a student of B.E. (Computer Science & Engineering), Technocrats Institute of Technology (TIT), Bhopal has successfully completed the Project Dissertation Stage on "**Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks"** under the Guidance of Asst. Prof. Mayank Pathak.

Asst. Prof. Mayank Pathak	**Dr. Vivek Sharma**	**Dr. Bhupendra Verma**
Project Guide	Head of Department	Director
(Computer Science & Engineering)	(Computer Science & Engineering)	(Computer Science & Engineering)
TIT Bhopal	TIT Bhopal	TIT Bhopal

Technocrats Institute Of Technology

Bhopal (M.P.)

CERTIFICATE OF APPROVAL

This is to certify that the Major Project titled **"Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks"** submitted by **Mr. Tejas Verghese [Roll No. 0111CS141111]**, has been examined by us and is hereby approved for the award of degree **"Bachelor of Engineering in Computer Science & Engineering"**, for which it has been submitted. It is understood that by this approval the undersigned so not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the dissertation only for the purpose for which it has been submitted.

(Internal Examiner)	**(External Examiner)**
Date:	**Date:**

ACKNOWLEDGEMENT

I am grateful to my guide **Asst. Prof. Mayank Pathak** for providing me the opportunity to work in an exciting and challenging field of Deep Learning. My interactions with him have been of immense help in defining my project goals and in identifying ways to achieve them. I am grateful to him for spending his valuable time to carefully review this Project Dissertation. And for his valuable insights and suggestions that have been very helpful in making this Project a success.

I am also grateful for all the help and support extended to me by **Prof. Kamlesh Chandravanshi**. A special thanks for helping me with the various aspects of the Project Report and Presentation.

I would like to thank **Dr. Vivek Sharma**, Head of the Computer Science & Engineering Department, for his constant support and encouragement.

I would like to show my gratitude for the kind words of encouragement and thank **Dr. Bhupendra Verma**, Director of the Computer Science & Engineering Department.

And also thank you to all my friends and colleagues for their valuable support and encouragement.

Tejas Verghese

[0111CS141111]

Key-Agricultural Commodities Daily Market Price Prediction Using Deep Neural Networks

ABSTRACT

Key-Agricultural Commodities Market Price Prediction using Deep Neural Networks is a project that focuses on data prediction for market price values of agricultural commodities. It uses agricultural commodity datasets from the Open Government Data Platform to train the neural network to predict future market price values, such as – Minimum price, Maximum Price and Modal Price.

Sayantan Bera, a writer at Livemint, has described the current agricultural sector crisis as such – Droughts, hailstorms and floods are becoming more frequent due to the effects of global warming. This has led to some serious problems in the agricultural sector. Most Indian farmers are not covered by insurance and all do not receive relief for crop damage. Yet these payouts averaged about Rs. 24,000 crore per year between 2014-15 and 2016-17. The total economic loss to agriculture could be many times higher - last year's Economic Survey noted that India incurs losses of about Rs. 62,000 crore annually in agriculture alone [1].

To provide a solution to Global Warming itself is beyond the scope of this project. However, we can suggest a means to solve the problem of the Economic loss in the agricultural sector. This project gives us the ability to predict which particular crops might incur losses and which ones might gain some substantial profits, with respect to the State, District, Market and Arrival Dates of these agricultural commodities. This knowledge might be useful to the Government, Insurance agencies and even to the farmers themselves, to reduce or at least manage their losses in the agricultural sector.

TABLE OF CONTENTS

Statement of Candidate…………………………………………………….. **I**

Certificate……………………………………………………………… **II**

Certificate of Approval…………………………………………………….. **III**

Acknowledgement………………………………………………………… **V**

Abstract………………………………………………………………… **VII**

List of Figures…………………………………………………………….. **XII**

1. INTRODUCTION AND BACKGROUND **2**

1.1. Introduction………………………………………………………….. **2**

1.2. Background…………………………………………………………… **3**

1.2.1. Neural Networks (NN)……………………………………………. **3**

1.2.2. Deep Learning Neural Networks (DNN)………………………… **4**

1.2.3. Activation Function………………………………………… **5**

1.2.4. Backpropagation…………………………………………………. **5**

1.2.5. Stochastic Gradient Descent…………………………………… **6**

2. LITERATURE SURVEY **8**

2.1. Existing Work…………………………………………………………. **8**

2.1.1. Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System……………………………………… **8**

2.1.2. Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications………………………………………………… **8**

2.1.3. Empirical Study on Agricultural Products Price Forecasting based on Internet-based Timely Price Information…………………………… **9**

2.1.4. Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression……………………………………… **9**

2.2. Deep Neural Networks…………………………………………………. **9**

2.3. Activation Function ………………………………………………….. **11**

2.4. Stochastic Gradient Descent……..…………………………………….. **13**

3. ANALYSIS AND DESIGN **16**

3.1. Requirement Analysis………………………………………………… **16**

3.1.1. Hardware Requirements…………………………………………… **16**

3.1.2. Software Requirements……………………………………………. **16**

3.2. Feasibility Analysis…………………………………………………… **17**

3.2.1. Economic Feasibility………………………………………………. **17**

3.2.2. Technical Feasibility……………………………………………….. **17**

3.2.3. Operational Feasibility…………………………………………… **18**

3.3. Project Design Approach…..………………………………………….. **19**

3.3.1. Class Diagram……………………………………………………….. **19**

3.3.2. Object Diagram……………………………………………………… **21**

3.3.3. Use Case Diagram…………………………………………………… **22**

3.3.4. Sequence Diagram…………………………………………………… **23**

3.3.5. E-R Diagram…………………………………………………………. **23**

3.3.6. Activity Diagram…………………………………………………… **24**

4. PROPOSED WORK AND IMPLEMENTATION **26**

4.1. Objective ……………………………………………………………… **26**

4.2. Motivation…………………………..…………………………………… **27**

4.3. Problem Statement……………………………………………………… **28**

4.4. Problem Definition……………………………………………………... **28**

4.5. Proposed Work………………………………………………………… **29**

4.6. Proposed Architecture………………………………………………….. **30**

4.7. Implementation Details…………………………………………………. **31**

4.7.1. Independent Variables……………………………………… **31**

4.7.2. Dependent Variables………………………………………… **32**

4.7.3. Accuracy & Precision……………………………………….. **33**

4.7.4. Scale………………………………………………………… **34**

4.7.5. Cost…………………………………………………………… **34**

4.7.6. Project Limitations…………………………………………… **35**

5. RESULT **37**

5.1. Login Window ………………………………………………………… **37**

5.2. Selector Window……………….……………………………………. **37**

5.3. Main Window……………………………………………………….. **38**

5.4. Dialog Windows……..………………………………………………. **39**

5.4.1. Sign-Up Window………………………………………………… **39**

5.4.2. Update (Row) Window…………………………………………… **39**

5.4.3. Delete (Row) Window……………………………………………. **40**

5.5. Prediction……………………………………………………………… **41**

6. CONCLUSION AND FUTURE WORKS **43**

6.1. Conclusion…………………………………………………………… **43**

6.2. Future Work………………………………………………………….. **44**

References……………………………………………………………….. **45**

LIST OF FIGURES

Figure 1.1	Major Project Logo……………………………………………	**2**
Figure 1.2	Biological vs. Artificial Neurons…………………………...	**4**
Figure 2.1	Deep Neural Network………………………………………	**10**
Figure 2.2	Threshold (Step) Function………………………………….	**11**
Figure 2.3	Sigmoid Function…………………………………………..	**12**
Figure 2.4	Rectifier Function…………………………………………..	**12**
Figure 2.5	Hyperbolic Tangent Function………………………………	**13**
Figure 3.1	Class Diagram……………………………………………...	**21**
Figure 3.2	Object Diagram……………………………………………	**22**
Figure 3.3	Use Case Diagram…………………………………………	**22**
Figure 3.4	Sequence Diagram………………………………………….	**23**
Figure 3.5	E-R Diagram……………………………………………….	**23**
Figure 3.6	Activity Diagram…………………………………………	**24**
Figure 4.1	Farming in India…………………………………………..	**26**
Figure 4.2	The Proposed Flow Diagram………………………………	**30**
Figure 4.3	Independent Variable DataFrame………………………….	**31**
Figure 4.4	Dependent Variable DataFrame……………………………	**32**
Figure 4.5	State of the Agricultural Sector in India……………………	**33**
Figure 5.1	The Login Window………………………………………..	**37**
Figure 5.2	The Selector Window……………………………………..	**38**
Figure 5.3	The Main Window………………………………………….	**38**
Figure 5.4	The Sign-Up Window………………………………………	**39**
Figure 5.5	The Update (Row) Window………………………………..	**40**
Figure 5.6	The Delete (Row) Window………………………………	**40**
Figure 5.7	The Prediction……………………………………………..	**41**

Chapter I

INTRODUCTION AND BACKGROUND

Chapter - I

INTRODUCTION & BACKGROUND

1.1. <u>Introduction</u> :

Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks is a tool for market price data prediction of key agricultural commodities. It allows a user to create a dataset of values that includes information such as State, District, Market, Commodity, and Arrival Date. This dataset of values can either be created one value at a time, or it can be imported into the software from a CSV file (with a .csv extension). Then the software uses the Deep Neural Network Model to predict the values for the Minimum, Maximum and Modal Prices for each record in the dataset.

Figure 1.1: Major Project Logo

This software is capable of predicting the minimum price, maximum price and the modal price of an agricultural commodity that has been brought to the market by a farmer on any particular date. This can help the government, or an insurance company in policy preparation to help farmers in case of crop damage. As the software uses past commodity market price data as a training set to predict future values, it is capable of making accurate predictions as per how much revenue can be potentially earned by a farmer of a particular state, district, market, and who has auctioned off their crop on a particular date.

Therefore, as this software makes predictions ahead of time about the potential revenue of a farmer who grows a particular commodity and then auctions it off at any market in any district of the state where the commodity is usually grown or auctioned. It opens for us the possibility of predicting the future market scene in the agricultural sector. Also it gives farmers the hope of better recompensation in case of crop damage. Governments will also be able to better manage the loss of revenue in agriculture, as they will now have be able to quantify the revenue lost in terms of the expected revenue of the harvest.

1.2. Background :

1.2.1. Neural Networks (NN) :

In a paper titled "Artificial Neural Networks as Models of Neural Information Processing" by Marcel van Gerven *et al.*, Artificial Neural Networks (ANNs) or Neural Networks are described as computing systems vaguely inspired by the biological neural networks that constitute animal brains [2]. Such systems learn or progressively improve their performance on tasks by considering examples, generally without task-specific programming. As described in the online book titled "Neural Networks and Deep Learning" by Michael Nielsen, they use programming modules called perceptrons which are artificial neurons that constitute the most basic form of a neural network. A perceptron accepts input values (x) and assigns weights (*w*) to it based on the training dataset. Then the sum of products of the input values and the assigned weights ($\sum w_i x_i$), is calculated and that in turn is run through an Activation function to product an output value denoted by y [3].

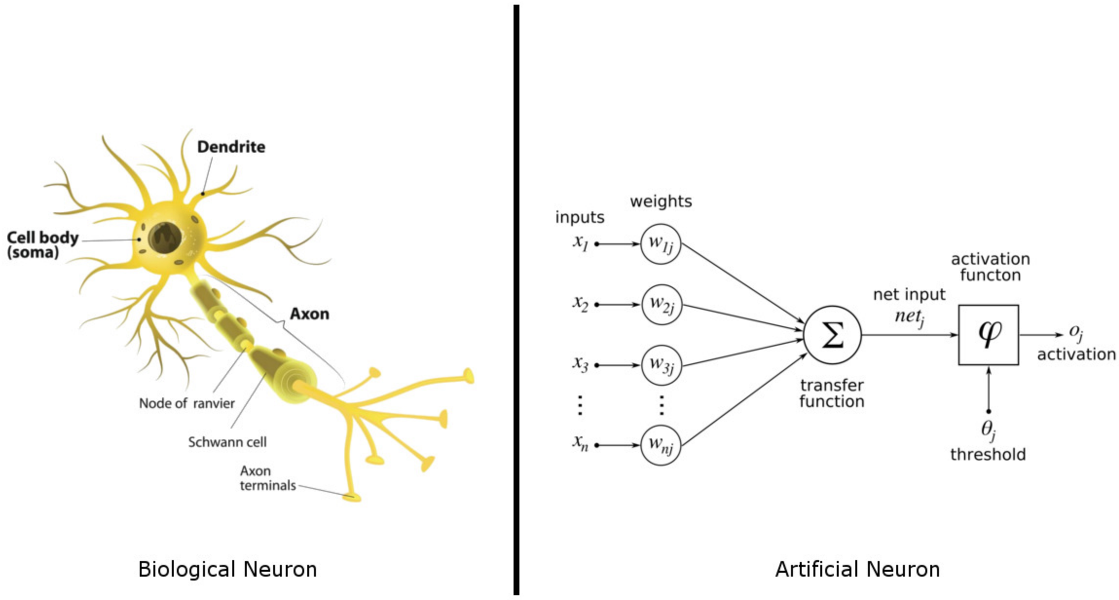


Figure 1.2: Biological vs. Artificial Neuron

1.2.2. Deep Learning Neural Networks (DNN) :

A Deep Neural Network (DNN) as defined by Bengio *et al.* in the paper "Learning Deep Architectures for AI", is an artificial neural network with 2 or more hidden layers between the input and output layers. DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives. The extra layers enable composition of features from lower layers, potentially modeling complex data with fewer units than a similarly performing shallow network [5]. DNNs are typically feedforward networks in which data flows from the input layer to the output layer without looping back.

Deep architectures include many variants of a few basic approaches. Each architecture has found success in specific domains. Some of its types are –

- Recurrent Neural Networks (RNNs), in which data can flow in any direction, are used for applications such as language modelling. Long short-term memory is particularly effective for this use.
- Convolutional Neural Networks (CNN) are used in computer vision. CNNs also have been applied to acoustic modeling for automatic speech recognition (ASR).

1.2.3. <u>Activation Function</u> :

Xavier Glorot *et al.* have in their paper titled "Deep sparse rectifier neural networks" described the activation function with respect to the role it plays in any neural network model [12]. To get an output value in any Artificial Neural Network we use the sum of products of inputs (x) and their corresponding Weights (*w*) and apply an Activation Function denoted by f(x) to it. This gives us the output of that layer and this output value is fed into the next layer as an input value.

Activation Functions are necessary for ANN based models as without this feature the model would be a simple linear function. And a linear function is a polynomial of one degree, i.e. it has limited complexity and also has only a limited ability for learning complex functional mappings from the data. An Activation Function is needed to achieve non-linearity in the model. A Non-linear function is a polynomial that has a degree that is more than one. Therefore, it is more suitable for complex data analysis.

1.2.4. <u>Backpropagation</u> :

As described by Yann LeCun *et al.* in the paper titled "Efficient Backprop", Backpropagation is short for "Backward

Propagation of Errors". It is an algorithm for supervised learning of artificial neural networks using gradient descent. Given an artificial neural network and an error function the method calculates the gradient of the error function with respect to the neural network's weights and then updates the weights to account for the error margin [4].

1.2.5. Stochastic Gradient Descent :

As mentioned in the online book "Neural Networks and Deep Learning" by Michael Nielsen, Stochastic gradient descent (often shortened to SGD), also known as incremental gradient descent, is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions [6]. In other words, SGD tries to find minima or maxima of a gradient by iteration.

In Stochastic Gradient Descent, you process each row of the dataset and update the weights for every iteration of the rows, until the entire dataset is completely processed by the neural network model. Surprisingly, Stochastic Gradient Descent is faster than Batch Gradient Descent, simply because SGD processes the dataset one row at a time whereas BGD will load the entire dataset onto memory at once.

Chapter II

LITERATURE SURVEY

Chapter – II

LITERATURE SURVEY

2.1. Existing Work :

2.1.1. Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System :

This system was proposed by Girish K. Jha *et al.* and published in the Agricultural Economics Research Review (AERR) [7]. They had proposed a system where forecasts of food prices are intended to be useful for farmers, policymakers and agri-business industries. They have used an Artificial Neural Network model for monthly wholesale price forecasting of soybean and rapeseed-mustard. This model uses a Time-Delay Neural Network with a single hidden layer, to predict the monthly wholesale prices of soybean and rapeseed-mustard using a time-series dataset as training set.

2.1.2. Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications :

According to the paper "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications" by Manpreet Kaur *et al.* [8], data mining is emerging as an important field of research in agriculture crop price analysis. The paper discusses the application and techniques of data mining in agriculture. It evaluates different data mining techniques, such as – K-Means, K-Nearest Neighbor (KNN), Artificial

Neural Networks (ANN), and Support Vector Machine (SVM), on different datasets.

2.1.3. Empirical Study on Agricultural Products Price Forecasting based on Internet-based Timely Price Information :

The research by Wang Xin *et al.* published through the paper “Empirical Study on Agricultural Products Price Forecasting based on Internet-based Timely Price Information” [9], suggests a method to use the Web-crawler technology to enable timely and accurate collection of agricultural products prices among 214 large-scale wholesale markets in China. The article suggests using the collected data to develop a quantitative analysis predictive model of wholesale prices.

2.1.4. Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression :

The paper “Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression” was presented by Andres M. Ticlavilca *et al.* at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market–Risk Management [10]. This paper suggests the usage of Multivariate Relevance Vector Machine (MVRVM) that is based on a Bayesian learning machine approach for regression. The paper uses Bootstrapping methodology to compare and analyze the robustness of the MVRVM to a multivariate ANN.

2.2. Deep Neural Networks :

A Deep Neural Network (DNN) is basically a deep learning neural network, this means that it is an advanced form of an Artificial Neural Network (ANN) which is capable of deep learning. Just like an Artificial Neural Network a

Deep Neural Network also has an Input Layer, Hidden Layer and an Output Layer. However, an ANN only has a single hidden layer whereas a DNN will have at least 2 hidden layers in its architecture.

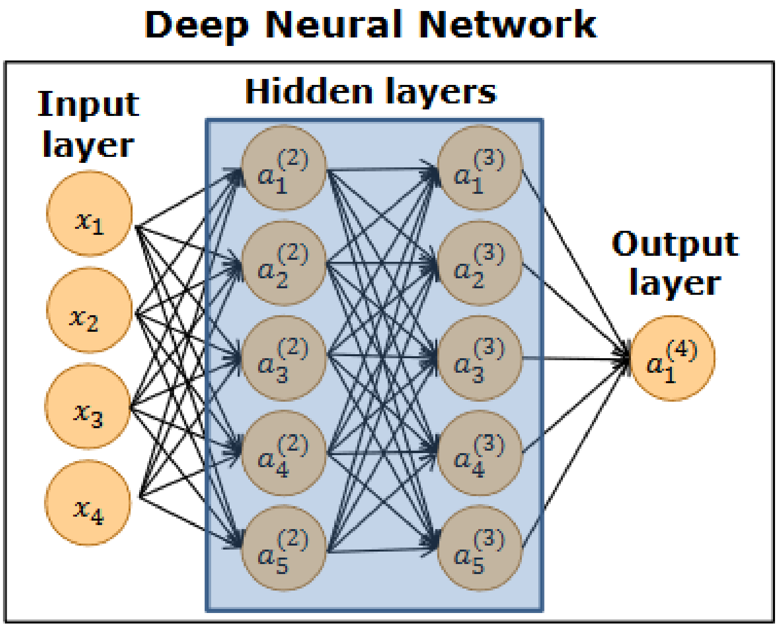


Figure 2.1: Deep Neural Network

The Figure 2.1 shows a DNN that has 1 input layer, 2 hidden layers and 1 output layer. The DNN works by taking in the input values one record at a time. As shown in the input layer - x_1, x_2, x_3 and x_4, are input values from a single record (or row of data in case of a dataset), i.e. the column values of each record in the dataset are loaded into the DNN's input layer. All these input values are given assigned a weight (w_i) and then the product of these input values (x_i) and respective weights (w_i) are fed into the first hidden layer which processes all these values and creates the most viable relations based on the training data. The output from the first hidden layer is fed into the second hidden layer which has the same job as the first hidden layer, however the relationships built between the first and second hidden layers will give the DNN an edge over the traditional ANN as the connections formed by the second hidden layer will add more complexity to the DNN model. This complexity is beneficial to us as it allows the model to learn the training data in a deeper and more complex manner. Thereby, allowing for better forecasting of data values. The output of the second hidden layer is then finally fed into the output layer which will provide us an output value denoted by y.

A Deep Neural Network (DNN) as defined by Bengio *et al.* in their paper titled "Learning Deep Architectures for AI" [5], is defined an artificial neural network with 2 or more hidden layers between the input and output layers. DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives. The extra layers enable composition of features from lower layers, potentially modeling complex data with fewer units than a similarly performing shallow network.

2.3. Activation Function :

In the paper titled "Deep sparse rectifier neural networks" the authors Xavier Glorot *et al.* [12], suggest that activation functions are general mathematical functions that may be linear or non-linear in nature. But in the case of Deep Neural Networks they are exclusively non-linear functions. These are denoted by f(x). Activation functions are a necessary part of any artificial neural network as these are responsible for the activation of the artificial neuron.

There are generally 4 types of activation functions used in Neural Networks, they are –

- Threshold Function :

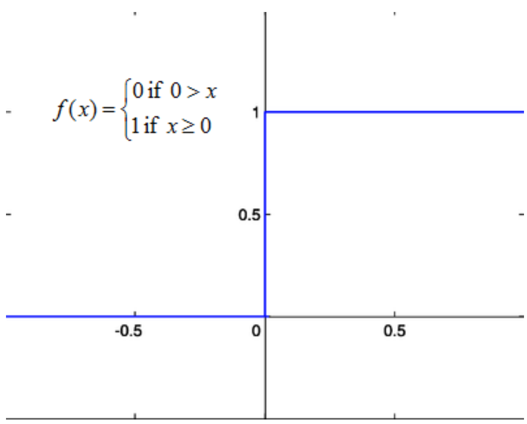
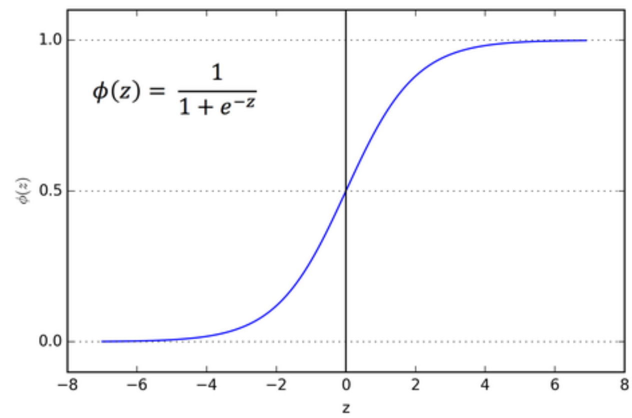


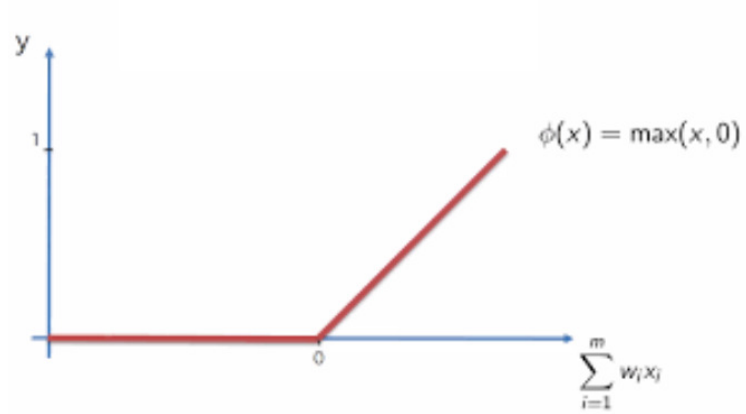
Figure 2.2: Threshold (Step) Function

- <u>Sigmoid Function</u> :



<u>Figure 2.3: Sigmoid Function</u>

- <u>Rectifier Function</u> :



<u>Figure 2.4: Rectifier Function</u>

- Hyperbolic Tangent Function :

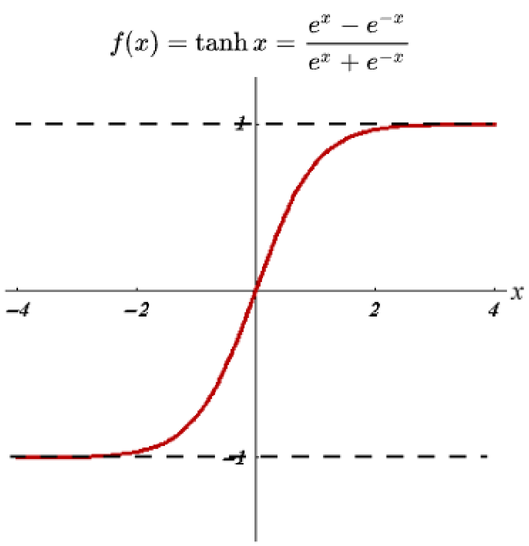


Figure 2.5: Hyperbolic Tangent (tanh) Function

In the scope of this project, we will be using the Hyperbolic Tangent (tanh) as our Activation Function in both of the hidden layers of our Deep Neural Network models.

2.4. Stochastic Gradient Descent :

Stochastic Gradient Descent, also known as Incremental Gradient Descent. According to Wikipedia, it is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions. In simple words, SGD is used to find minima and maxima by iteration.

As mentioned in the online book “Neural Networks and Deep Learning” by Michael Nielsen, Stochastic gradient descent (often shortened to SGD), also known as incremental gradient descent, is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions [6]. In other words, SGD tries to find minima or maxima of a gradient by iteration.

In Stochastic Gradient Descent, you process each row of the dataset and update the weights for every iteration of the rows, until the entire dataset is completely processed by the neural network model. Surprisingly, Stochastic Gradient Descent is faster than Batch Gradient Descent, simply because SGD processes the dataset one row at a time whereas BGD will load the entire dataset onto memory at once.

Chapter III

ANALYSIS AND DESIGN

Chapter – III

ANALYSIS & DESIGN

3.1. Requirement Analysis :

3.1.1. Hardware Requirements :

The following are the basic hardware requirements for this project –

- Memory: 8GB or more of System RAM.
- Hard Drive: 25GB or more of free Hard Drive Space.
- Processor: any intel or amd dual-core processor with at least 2GHz or better.

3.1.2. Software Requirements :

The following are the basic software requirements for the project –

- OS Platform: Linux 4.10 or newer (Recommended), MacOS, Windows 8.1 or newer.
- Python 3.5 or newer.
- NumPy, for computations.
- Pandas, for dataset manipulation.
- PyQt5 for the GUI Interface.
- Scikit-learn, for computations and data preprocessing.
- Theano libraries, as it is a dependency for Keras.
- TensorFlow libraries, also a dependency for Keras and necessary for neural networks and related computations.
- Keras libraries, for fast modeling of the neural networks.

3.2. **Feasibility Analysis :**

3.2.1. Economic Feasibility :

Since this is a Deep Learning Project it will be highly economical in its nature as it does not require a network connection or bandwidth. All its training data needs are fulfilled by the Training datasets from the Open Government Data Platform where it is available free for all to use.

The software will follow the freeware software standards. Bug fixes and maintenance might have some associated cost. At the initial stage the potential customers will be the government's agriculture related departments, crop insurance agencies, agriculture associated businesses and farmers.

Aside from the cost associated with maintaining the software there will be many real time benefits to farmers, governments, crop insurers, and agricultural businesses as the project forecasts the potential future market scenario in the agricultural sector. It will predict values such as the minimum, maximum and modal prices of key-agricultural commodities. These predictions are based on the past market trends in the Indian agricultural sector and therefore will be extremely useful to those associated with the agricultural sector in India.

From these it is clear that the project "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" is economically feasible.

3.2.2. Technical Feasibility :

The project "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" is a software project that is based on Deep Neural Networks. The main technologies and tools that are associated with Deep Learning Neural Networks are –

- Python
- NumPy
- Pandas
- Scikit-learn
- TensorFlow
- Theano
- Keras
- PyQt5, for GUI interface

Each of these technologies are freely available and the technical skills required are easily attainable. Time limitations of the project development and the ease of implementation using these technologies are synchronized.

Initially, the Deep Neural Network will be trained using free datasets from the Open Government Data Platform. However, if the project gains significant approval from the government and agriculture associated businesses then the datasets to train the model can be acquired live and therefore data can be captured in greater detail. This will be highly useful in making the project's prediction capabilities more accurate.

Therefore, from these it is clear that the project "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" is technically feasible.

3.2.3. Operational Feasibility :

As "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" is a Deep Learning Neural Networks based project, it only requires one to enter the dataset for which data prediction is needed. Therefore, it's operation is fairly simple and requires less effort from the user.

However, with time the models that have been trained will need to be maintained consistently. As Global Warming is affecting agriculture in India on a considerably large scale, there is a need to constantly train the model with new data. As we all know that these days every new year instantly becomes the hottest year yet. We can still make accurate predictions but just to ensure that the effects of global warming do not have any unwanted effects on our data prediction capability we must ensure consistent retraining of the models associated with the project. But other than that the project continues to remain useful to those in the agricultural sector of India.

From these it is clear that the project “Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks” is operationally feasible.

3.3. Project Design Approach :

3.3.1. Class Diagram :

The Class Diagram describes the various classes, data members and member functions of the project. The various classes are described as follows –

- Ui_LoginWindow Class :

  This class is responsible for accepting the input values for username and password and then check in the login database’s user table if the user is registered. If yes, then it will allow the user to login else it won’t.

  The variables associated with this class are – userEdit and passEdit.

  The functions associated with this class are – setupUi(), showSelectorWindow(), showLoginWindow(), loginApplication() and closeApplication().

- Ui_SelectorWindow Class :

  This class is responsible for providing the user a selection of 12 commodities. Whichever one the user selects is kept as the default commodity to make predictions for.

  There is only one associated variable, i.e. commodity.

  The functions of this class are – setupUi(), showSignUp(), showMainWindow(), closeApplication().

- Ui_MainWindow Class :

  This class is responsible for all the interactions between the user and the Deep Neural Network models in the project.

  The variables of this class are – commodity, stateChoice, districtChoice, marketChoice and dateEdit.

  The functions of this class are – setupUi(), choiceState(), choiceDistrict(), choiceMarket(), showUpdateDialog(), showDeleteDialog(), showSignUp(), showSelectorWindow() and closeApplication().

- KAC_Predictor Class :

  This class is responsible for loading up the 3 models for prediction of values – KAC_MinModel.h5, KAC_MaxModel.h5 and KAC_ModModel.h5, and then use those models to make predictions and save those predictions to the combo_loader database's output table.

  The variables of this class are – dataset, X, y_min_pred, y_max_pred and y_mod_pred.

  There are only 2 functions in this class – loadModels() and predictor().

- Ui_SignUpDialog Class :

  This class is responsible for allowing new users to sign-up to use the software.

- Ui_UpdateDialog Class :

  This class is responsible for updating the input dataset if and when required by the user.

- Ui_DeleteDialog Class :

  This class is responsible for deleting or removing any records that the user needs removed from the input dataset.

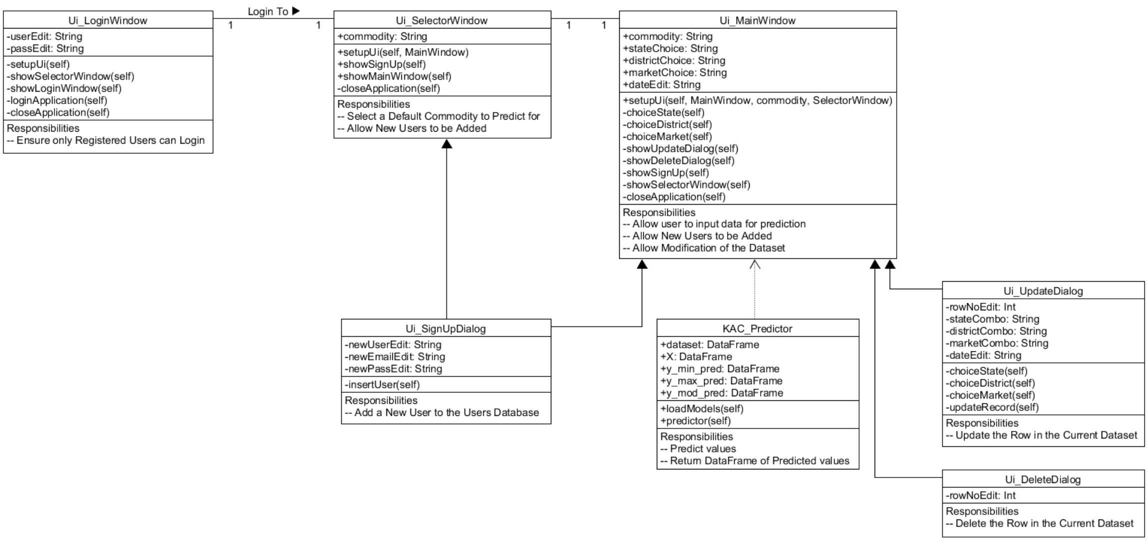


Figure 3.1: Class Diagram

3.3.2. Object Diagram :

This Object diagram shows the various objects along with their attributes that are created in the project as it is used. The following are the objects and their respective classes –

- ui : Ui_LoginWindow
- ui : Ui_SelectorWindow

- ui : Ui_MainWindow
- predictor : KAC_Predictor
- ui : Ui_SignUpDialog
- ui : Ui_UpdateDialog
- ui : Ui_DeleteDialog

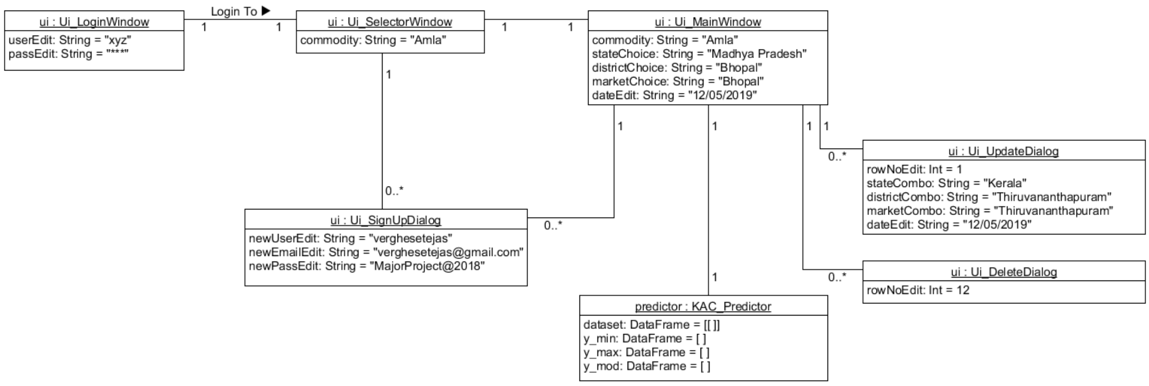


Figure 3.2: Object Diagram

3.3.3. Use Case Diagram :

The following figure shows the Use-Case Diagram of the project. It helps us understand the flow in which it is meant to be used by any potential user of the software.

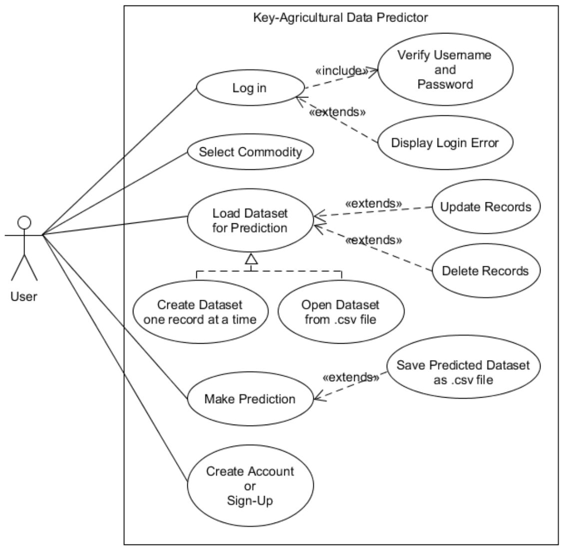


Figure 3.3: Use Case Diagram

3.3.4. Sequence Diagram :

The Sequence Diagram shows us the sequence of the interactions between the various objects and classes of the program.

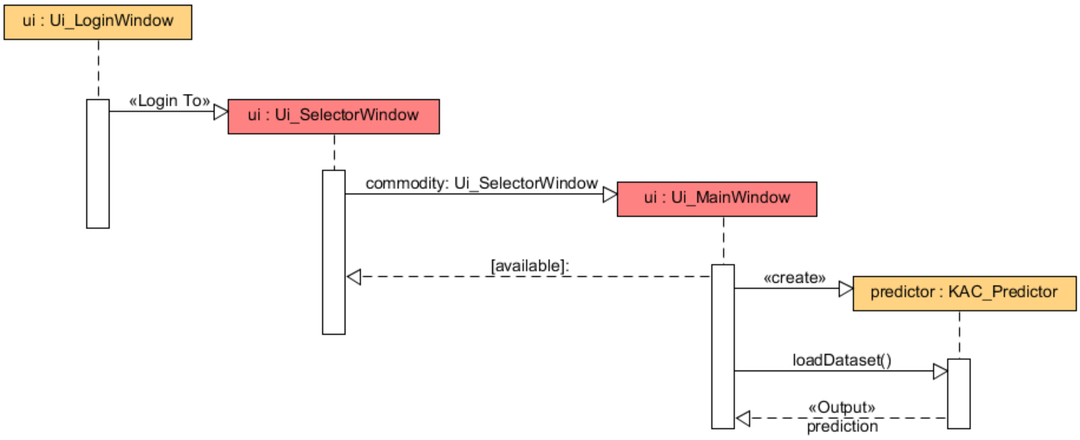


Figure 3.4: Sequence Diagram

3.3.5. E-R Diagram :

The Entity-Relationship Diagram tells us about the interaction if any between the tables that are used in the scope of the project.

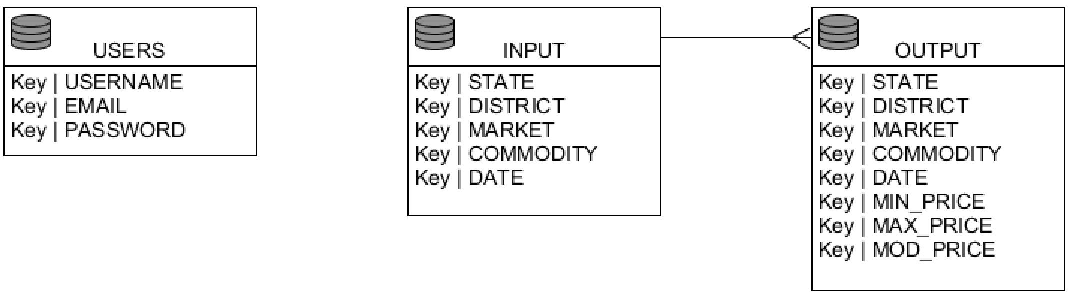


Figure 3.5: E-R Diagram

3.3.6. Activity Diagram :

The Activity Diagram tells us about the activity flow of the software system. As the designated user utilizes the software, the instances and parts of the software are shown as they are encountered by the user. It also defines a start point for the software and an end point.

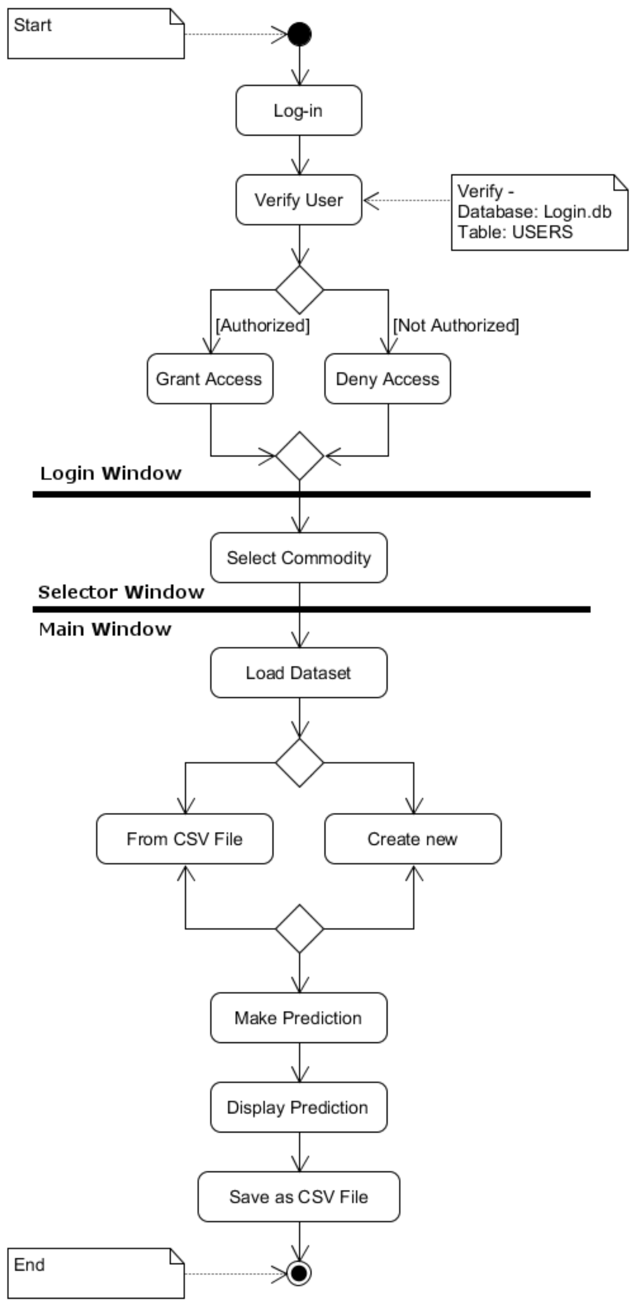


Figure 3.6: Activity Diagram

Chapter IV

PROPOSED WORK

AND

IMPLEMENTATION

Chapter – IV

PROPOSED WORK & IMPLEMENTATION

4.1. Objective :

The main objective of this application is to predict the future market scene for the use of the government, crop insurers, and farmers. The Agricultural Data Predictor is a software tool capable of Predicting the market price data associated with certain Key-Agricultural Commodities. The software is intended to be used by government's agricultural departments, crop insurance agencies or even the farmers themselves.

Figure 4.1: Farming in India [11]

It will help them decide and plan for any possible losses in revenue that the farmer might have to incur during the auctioning of the crops. These losses may be caused due to various reasons, such as – crop damage, lack of quality of the crop, storage related issues and likewise. If the government and other interested parties can have some prior knowledge of the possible revenue that might be earned by the farmer, then it is the farmer who will ultimately benefit from this as the government and crop insurers now know how much the farmer can potentially earn.

4.2. <u>Motivation</u> :

One of the major motivators for this project is the current agricultural crisis. As farmer suicide rates are increasing and there is a need to innovate and re-invent farming for the sake of those whose very lives depend upon it. We must find new ways to make farming a profitable business if we ever hope to help those who need our support in the agricultural sector.

The following is an excerpt from a Blogspot article titled "An 11-point agenda for resurrecting Indian agriculture and restoring the pride in farming" by Devinder Sharma, Indian Author and Food Policy Analyst –

"Indian agriculture is faced with a terrible agrarian crisis. It is a crisis primarily of sustainability and economic viability. The severity of the crisis can be gauged from the spate of farm suicides. In the past 17 years, close to 3 lakh farmers reeling under mounting have preferred to commit suicide. Another 42 per cent want to quit agriculture if given a choice. The spate of farmer suicide and the willingness of farmers to quit agriculture is a stark reminder of the grim crisis." [11]

The ability to avoid some problems that we might face in the agricultural sector might be possible if we can know or predict the future market scenario using the past trends in the agricultural markets.

The following is another excerpt from a Livemint article titled "How vulnerable is Indian agriculture to climate change" by Sayantan Bera, Writer at Livemint –

"*Climate change has hit farmers and governments hard as drought, hailstorms and floods become more frequent. Most Indian farmers are not covered by insurance and all do not receive relief for crop damage."* [1]

We can create a better way to recompensate farmers for their losses. If we know the potential revenue based on past agricultural market trends then the software is capable of using that to predict the future scenario of the same market. This may not be enough to finish all our problems in the agricultural sector, however it can help us better manage and regulate the agricultural market for the benefit of the average Indian farmer.

4.3. **Problem Statement** :

To create a software application that is capable of making accurate market value predictions for Key-Agricultural Commodities in the Indian Agricultural Sector using data about the State, District, Market, Commodity and Arrival Date.

4.4. **Problem Definition** :

The Project titled "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" focuses on predicting the future prices of key agricultural commodities, such as - amla, blackgrams, chickpeas white, cummin seeds, dry peas, ladiesfinger, lemon, lentils, mint, paddy, persimmon and raddish. It is intended to predict the minimum, maximum and modal prices of these commodities.

It uses Deep Learning Neural Networks for data prediction and regression. The project uses datasets from the Open Government Data Platform as a training set of Independent and Dependent Variables to predict the market prices of the commodities in the near future. This prediction can be used to gain prior knowledge about the future market values of agricultural commodities and then we can use this information to manage the agricultural sector in a better way, where any farmers who could lose revenue due to crop damage or any such issues can be properly and accurately compensated using the market price predictions for that particular farmer's agricultural commodity. It will help banks, crop insurers and farmers as well to know their profit margins or even potential loss margins ahead of time. And therefore, they can prepare for any potential loss for even better balance their entire system of agricultural loss management.

Therefore, this project opens a new avenue for the agricultural sector where the government, the banks, the crop insurance agencies and ultimately, the farmers will all benefit simply by knowing the potential of their agricultural earnings.

4.5. Proposed Work :

The Project "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" is a Data Prediction Project based on Deep Neural Networks. In this project, we propose the use of Agricultural market price datasets from the Open Government Data Platform (https://data.gov.in/) to forecast the future agricultural market scenario.

The project proposes the use of data from columns such as – State, District, Market, Commodity, Arrival Date to predict values for Min_Price, Max_Price and Mod_Price. On the Open Government Data Platform the Indian Government organizations will consistently share data related to the functioning of the various governmental organizations. This data is accessible to the general public under the Open Data initiative of the Government of India.

These predicted values can be of great use to the various organizations working in the Indian agricultural sector. This project gives them the ability to forecast the future of the agricultural markets all across the country. We are proposing 3 Deep Neural Network models with 2 hidden layers each to predict market values of agricultural commodity prices in India. Of the 3 DNN models each will be responsible for – Minimum Prices Prediction, Maximum Prices Prediction and Modal Prices Prediction, respectively. All the values are predicted simultaneously to ensure uniformity. The models will have 5 input neurons each, followed by 5 neurons in the first hidden layer, followed by another 3 neurons in the second hidden layer and finally a single neuron in the output layer. The models will each be trained for a 100 epochs with an appropriate batch size, instead of the whole dataset at once. After their training the models are saved along with the various weights and measures associated with it.

Therefore, our proposed plan allows us to use 3 DNN models to predict 3 very important variables in the dataset in the future – i.e. the future prices of agricultural commodities in the country of India.

4.6. Proposed Architecture :

The Activity Diagram tells us about the activity flow of the software system. As the designated user utilizes the software, the instances and parts of the software are shown as they are encountered by the user. It also defines a start point for the software and an end point.

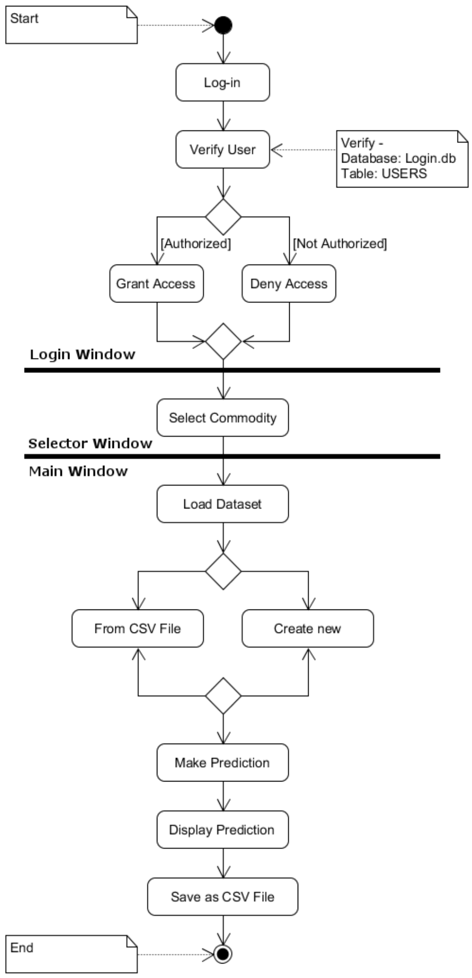


Figure 4.2: The Proposed Flow Diagram

4.7. <u>Implementation Details</u> :

4.7.1. <u>Independent Variables</u> :

In any machine learning model, including this Deep Neural Network, independent variables are the values that are used to predict the dependent variables. These values might include categorical data, floating point data, binary data, or even integral data.

X - DataFrame

Index	0	1	2	3	4
0	Andhra Pradesh	Anantapur	P. Ramacharyulu…	Amla	733041
1	Chattisgarh	Bastar	Gidam	Amla	733485
2	Chattisgarh	Bastar	Gidam	Amla	733486
3	Chattisgarh	Bastar	Gidam	Amla	733489
4	Chattisgarh	Bastar	Gidam	Amla	733490
5	Chattisgarh	Bastar	Narayanpur	Amla	732591
6	Chattisgarh	Bastar	Narayanpur	Amla	732598
7	Chattisgarh	Bastar	Narayanpur	Amla	732605
8	Chattisgarh	Bastar	Narayanpur	Amla	732606
9	Chattisgarh	Bastar	Narayanpur	Amla	732614
10	Chattisgarh	Bastar	Narayanpur	Amla	732619
11	Chattisgarh	Bastar	Narayanpur	Amla	732625
12	Chattisgarh	Bastar	Narayanpur	Amla	732626
13	Chattisgarh	Bastar	Narayanpur	Amla	732628
14	Chattisgarh	Bastar	Narayanpur	Amla	732630

Format | Resize | Background color | Column min/max | OK | Cancel

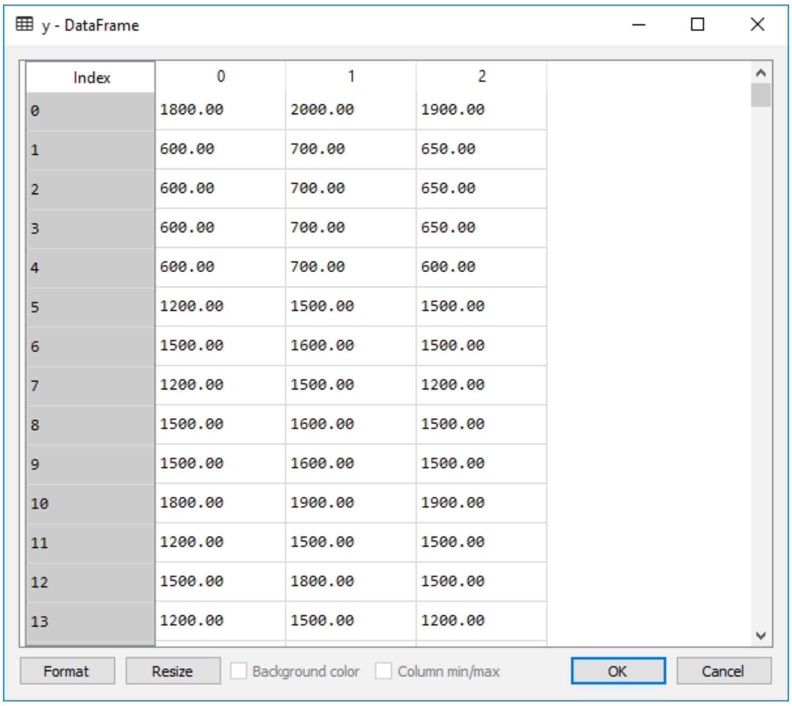
<u>Figure 4.3 : Independent Variables DataFrame</u>

In the scope of this project, the following columns are taken as independent variables – State, District, Market, Commodity and Arrival Date. This is done because the values in these columns do not depend upon the values in the dependent variables.

Therefore, it is possible for us to make accurate predictions about the dependent variables by using these independent variables as a training set for our Deep Neural Network.

4.7.2. <u>Dependent Variables</u> :

Dependent variables are the values that are to be predicted in the scope of the Neural Network Model. The Independent variables are used as a measure to predict the values for the dependent values. Now, in this project all three of our dependent variables are floating point numbers, as we will be making predictions for Price data.



Index	0	1	2
0	1800.00	2000.00	1900.00
1	600.00	700.00	650.00
2	600.00	700.00	650.00
3	600.00	700.00	650.00
4	600.00	700.00	600.00
5	1200.00	1500.00	1500.00
6	1500.00	1600.00	1500.00
7	1200.00	1500.00	1200.00
8	1500.00	1600.00	1500.00
9	1500.00	1600.00	1500.00
10	1800.00	1900.00	1900.00
11	1200.00	1500.00	1500.00
12	1500.00	1800.00	1500.00
13	1200.00	1500.00	1200.00

<u>Figure 4.4 : Dependent Variables DataFrame</u>

In the scope of this project, the following three columns are our dependent variables – Min_Price, Max_Price and Modal_Price. These variables are not all predicted by the same model. Instead, there are three models dedicated to predicting the Minimum Price, Maximum Price, and the Modal Price of the input dataset.

Therefore, we should remember that when we input a dataset and expect a prediction from the software, we are getting three predictions from three different but identical models of Deep Neural Networks.

4.7.3. Accuracy and Precision :

The Predictions the Deep Neural Networks make are not the exact float values as this is not a classification problem. It is a Regression problem and by that the accuracy is not exact. However, the accuracy of the current model after about 100 epochs is at a good enough levels for our project as the mean squared error (MSE) is shown to be at 0.01%, this means that the project is capable of making acceptable predictions for the Minimum Price, Maximum Price and Modal Price of individual agricultural commodities.

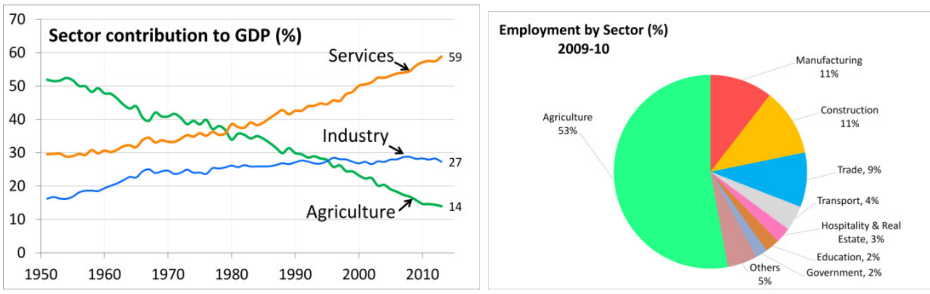


Figure 4.5 : State of the Agricultural Sector in India [13][14]

The precision of the model's predictions can be measured by tallying the predicted prices to the actual prices of the commodities that we are making the prediction for.

However, these Deep Learning Neural Networks are capable of increasing their own accuracy and precision using backpropagation with stochastic gradient descent. Therefore, it is not possible to say that the current accuracy levels will be the absolute accuracy of the model as the level of precision is bound to increase with respect to time. As the software is used to make regular predictions the backpropagation of errors is bound to increase its prediction capabilities.

4.7.4. Scale :

Key-Agricultural Data Predictor is a tool for market price data prediction of agricultural commodities, the scope or scale of the project is limited to the agricultural sector of the Republic of India.

The datasets used for training the Deep Neural Network models are open and free datasets that are made available to the general public by the Indian Government's Open Government Data Platform. Therefore, it is not possible for the models to predict market prices for commodities that are grow, bought, or auctioned outside of India and its Union Territories.

4.7.5. Cost :

The overall cost of the software project is very cheap and easily affordable. As the various aspects used in the making of this project are either free or Open-Source, thereby giving us the ability to create this project by simply investing time and knowledge.

The datasets used in the training of the Deep Neural Networks are open and Free for all to use. These are made available from the Open Government Data Platform of the Government of India.

The Programming language used for coding the software is Python 3.6, it is easy to download and install. Also free for all to use.

The various databases that were created for use in the Project are all SQLite 3 Databases, also free and open-source for all.

The library used for deep learning, TensorFlow, is an Open-Source library which is free and available to all. It was originally created by the Brain Team at Google. It is used for complicated computations related to Deep Learning Neural Networks.

Finally, the library used to implement the Deep Neural Network models, Keras, is another Open-Source library for fast implementation of Deep Learning Models. It is also free and open for all to use.

4.8. Project Limitations :

The Project does have a few limitations, these include –

- The scope of the predictions is limited to the Indian Agricultural Sector.
- The current models are only capable of making predictions for 12 key-agricultural commodities, including – amla, blackgrams, chickpeas white, cummin seeds, dry peas, ladiesfinger, lemon, lentils, mint, paddy, persimmon and raddish.
- The current model is also unfortunately incapable of distinguishing between the various sub-categories of the agricultural commodities.
- It does not account for weather events and instead it relies upon past records of market values.
- The level of security provided by the Login Page is not enough for real world situations.

Chapter V

RESULT

Chapter – V

RESULT

5.1. Login Window :

The Login Window will be responsible for providing a layer of security to the software. As the software makes predictions that might be crucial for the government, agriculture-related businesses or even directly to the farmers, the login window provides some much needed security as it allows only registered users to use the software in any capacity.

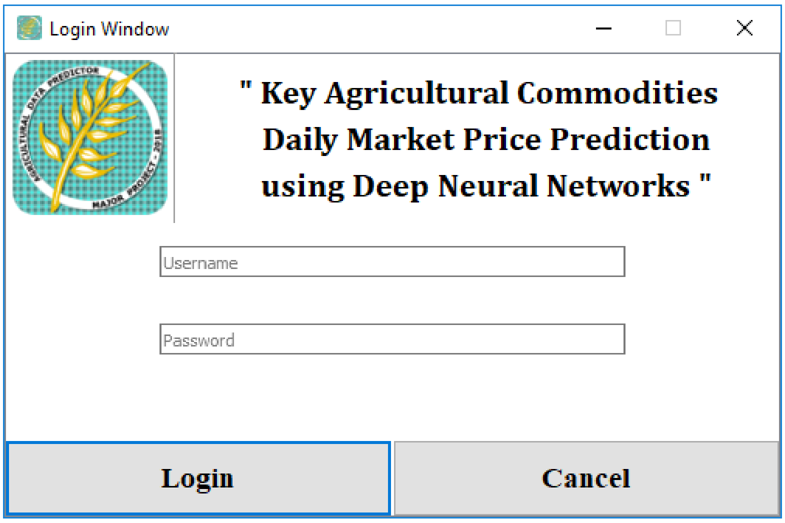


Figure 5.1: The Login Window

5.2. Selector Window :

The Selector Window shows the user a range of selections for agricultural commodities. Currently, the Deep Neural Network has been trained to make predictions for 12 key-agricultural commodities – amla, black grams, chickpeas white, cummin seeds, dry peas, ladies finger, lemon, lentils, mint, paddy, persimmon and raddish. Any of these commodities can be selected to move to the next window. This window also allows us to sign-up as a new user.

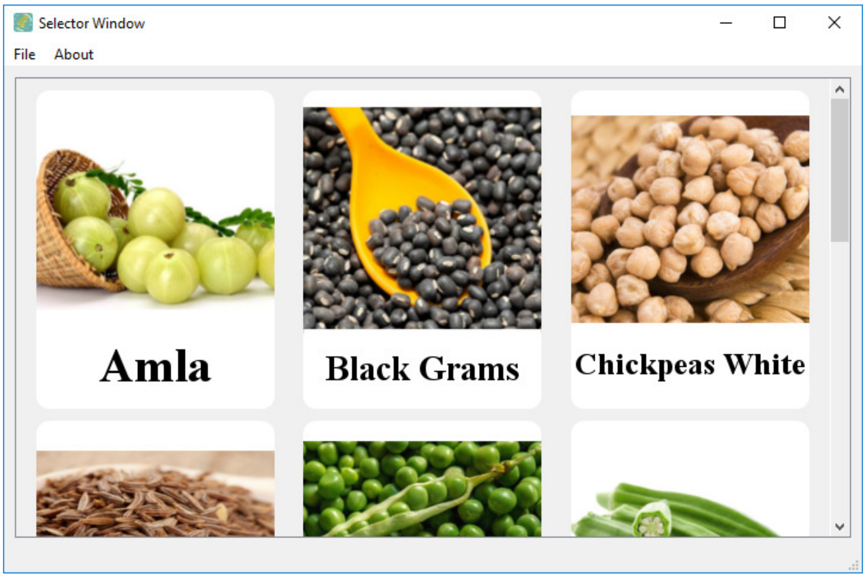


Figure 5.2: The Selector Window

5.3. Main Window :

The Main Window is the main interface of the project. It allows us to interact with the Deep Neural Network, to make predictions. It also allows us to create an input dataset, view the items in it and then use the DNN to make predictions for our dataset. We can either create new records one at a time or just load a dataset from a CSV file (*.csv). The main window also gives us an update window and a delete window to edit our dataset before making predictions for it.

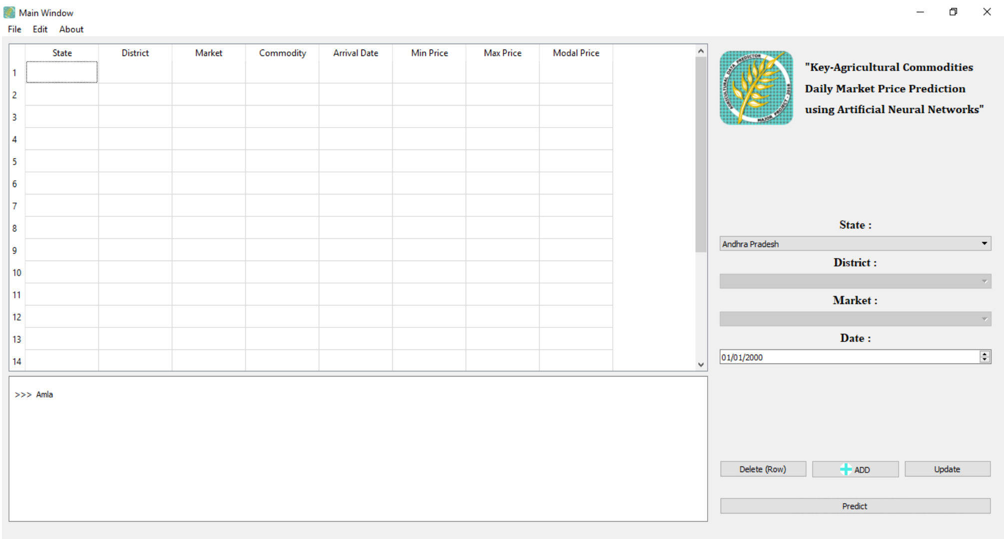


Figure 5.3: The Main Window

5.4. Dialog Windows :

These are the few Dialog Windows that are used for predefined tasks such as – to sign-up as a new user, to update a row, to delete a row, etc.

5.4.1. Sign-Up Window :

The Sign-up Window is a Dialog Window used to enter a new record into the User table in the login database. It allows us to create new users and share access to the software. It is available from the Selector Window and the Main Window.

Figure 5.4: The Sign-Up Window

5.4.2. Update (Row) Window :

The Update (Row) Window as the name suggests is a Dialog Window that allows users to update or modify the records of any row in the input dataset after it has been loaded into the software.

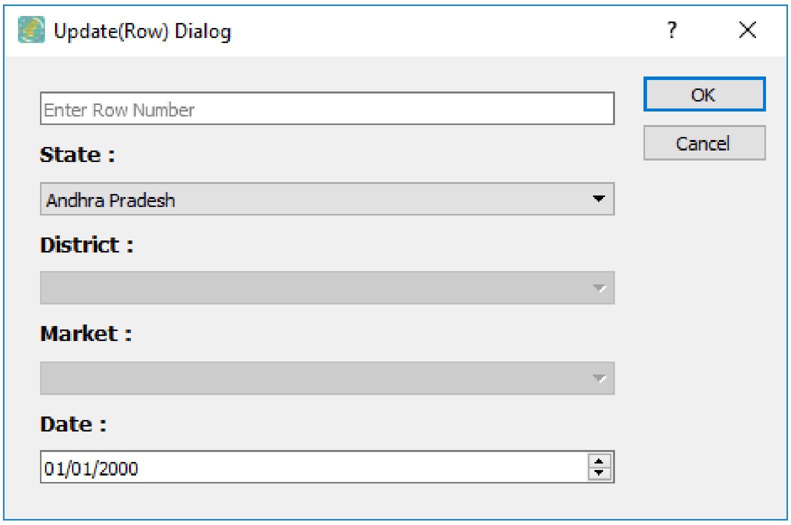


Figure 5.5: The Update (Row) Window

5.4.3. Delete (Row) Window :

The Delete (Row) Window as the name suggests allows the user to remove or delete any records from the dataset. It only accepts one parameter, i.e. row number of the record to be deleted or removed.

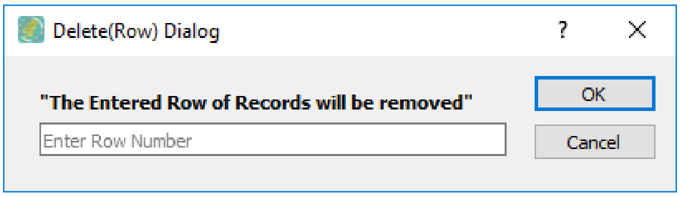


Figure 5.6: The Delete (Row) Window

5.5. **Prediction :**

The following figure shows the software as it displays the predicted values in the table widget that is inbuilt in the software's main window.

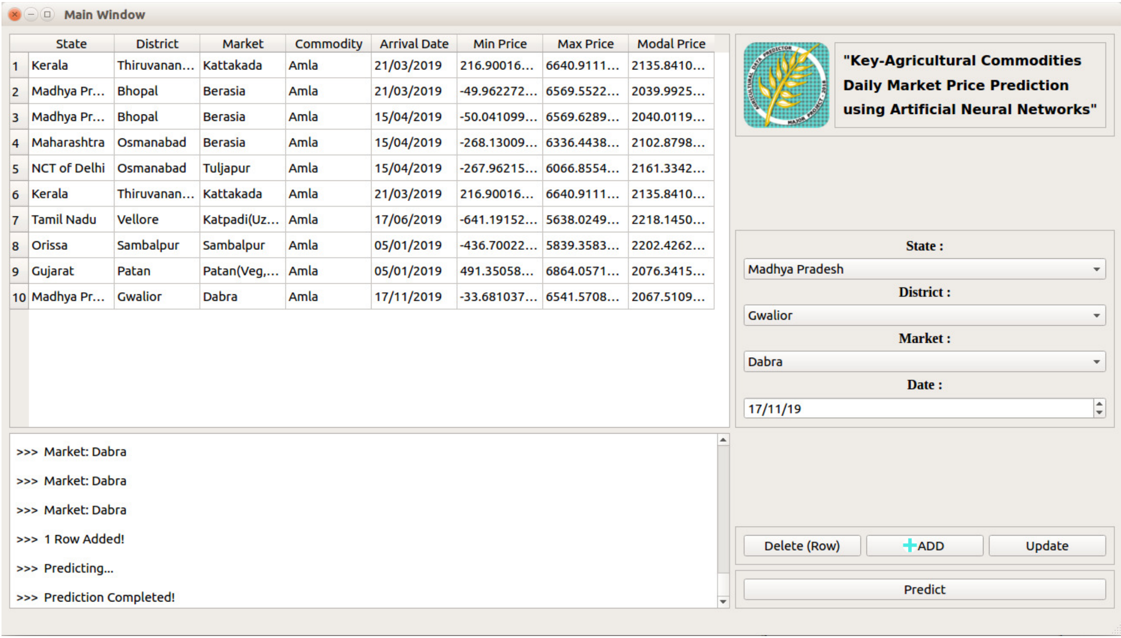


Figure 5.7: The Prediction

Chapter VI

CONCLUSION

AND

FUTURE WORKS

Chapter – VI

CONCLUSION & FUTURE WORKS

6.1. <u>Conclusion</u> :

In conclusion, the "Key-Agricultural Commodities Daily Market Price Prediction using Deep Neural Networks" project is capable of –

- The project uses 3 different model for – Minimum price prediction, Maximum price prediction and Modal price prediction, respectively.
- The models have a Mean Squared Error of 0.001%. This means that the loss in training is almost negligible. Therefore, making the prediction capabilities of the models highly accurate.
- It has a Login Window this allows only registered users to use the software. It also provides as an added layer of security for the project.
- The project is currently capable of making predictions for 12 key-agricultural commodities – amla, black grams, chickpeas white, cummin seeds, dry peas, ladies finger, lemon, lentils, mint, paddy, persimmon and raddish.
- The Selector Window will let a user select any one of the 12 commodities, this is because when making any sort of predictions in the professional work environment the forecasting will be done for one commodity at a time.
- In case the user wants to predict for multiple commodities there is the option to add more to the dataset.
- The Main Window will show display to the user the input dataset, and then display the output – the Minimum, Maximum and Modal price predictions.

- The Minimum price prediction will let the user know the minimum amount that can be earned by the farmer for any commodity which is auctioned at the selected state, district, market and date.
- The Maximum price prediction will let the user know the maximum amount that can be earned by the farmer for any commodity with the given parameters.
- The Modal price prediction will let the user know the most frequent amount that was earned by the farmer for any commodity under the given parameters.
- The project can also save the predictions that are made by it for future reference, as a CSV (*.csv) file.

6.2. <u>Future Works</u> :

In the near future, more capabilities can be added to this project and certain abilities that it already has can be made more accurate and precise. The following capabilities can be added or modified in this model –

- The ability to predict weather events and its effects on the crop production and its market price.
- The ability to predict over previously made predictions.
- Internet connectivity can be added to the model. It can use the internet connectivity to directly download new datasets from the Open Government Data Platform to re-train itself.
- The ability to share the prediction data over the internet.
- The data predictions can be more accurate, this will happen automatically with respect to time as the system uses the backward propagation of errors to correct itself.

References

[1] Sayantan Bera, “How vulnerable is Indian agriculture to climate change”, https://www.livemint.com/Politics/rxP2bO5krIzT48xMKIaGXO/How-vulnerable-is-Indian-agriculture-to-climate-change.html, 28 Mar. 2018.

[2] Marcel van Gervan *et al.*, “Artificial Neural Networks as Models of Neural Information Processing”, Front. Comput. Neurosci. Doi, 19 Dec. 2017.

[3] Michael Nielsen, “Neural Networks and Deep Learning”, http://neuralnetworksanddeeplearning.com/chap1.html#perceptrons, Dec. 2017.

[4] Yann LeCun *et al.*, “Efficient BackProp”, 1998.

[5] Bengio and Yoshua, "Learning Deep Architectures for AI”, Foundations and Trends in Machine Learning, 2009.

[6] Michael Nielsen, “Neural Networks and Deep Learning”, http://neuralnetworksanddeeplearning.com/chap2.html, Dec. 2017.

[7] Girish K. Jha and Kanchan Sinha, “Agricultural Price Forecasting Using Neural Network Model: An Innovative Information Delivery System”, Agricultural Economics Research Review, Vol.26 (No.2), pp. 229-239, Jul.-Dec. 2013.

[8] Manpreet Kaur, Heena Gulati and Harish Kundra, “Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications”, International Journal of Computer Applications(0975-8887), Vol.99 (No.12), Aug. 2014.

[9] Wang Xin and Wang Can, “Empirical Study on Agricultural Products Price Forecasting based on Internet-based Timely Price Information”, International Journal of Advanced Science and Technology, Vol.87, pp. 31-36, 2016.

[10] Andres M. Ticlavilca, Dillon M. Feuz and Mac McKee, “Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression”, NCCC-134 Conference, April 2010.

[11] Devinder Sharma, “An 11-point agenda for resurrecting Indian agriculture and restoring the pride in farming”, http://devinder-sharma.blogspot.in/2014/05/my-11-point-agenda-for-resurrecting.html, May 2014.

[12] Xavier Glorot *et al.*, “Deep sparse rectifier neural networks”, http://jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf, 2011.

[13] M Tracy Hunter, “1951 to 2013 Trend Chart of Sector Share of Total GDP for each year, India.png”, https://en.wikipedia.org/wiki/Economy_of_India#/media/File:1951_to_2013_Trend_Chart_of_Sector_Share_of_Total_GDP_for_each_year,_India.png, 18 Jul. 2014.

[14] M Tracy Hunter, “2010 Percent labor employment in India by its economic sectors.png”, https://en.wikipedia.org/wiki/File:2010_Percent_labor_employment_in_India_by_its_economic_sectors.png, 18 Jul. 2014.