Visvesvaraya Technological University,
Belgaum, Karnataka

Project Report
on

CROP PRICE PREDICTION

Submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering
in
Computer Science and Engineering

Submitted by

2SD16CS052 Mahesh Mandakki
2SD16CS131 Vadiraj Raichur
2SD17CS407 Neha Torgal

Under the Guidance of
Prof. Indira Umarji

Department of Computer Science and Engineering
S.D.M. College of Engineering and Technology
Dharwad, Karnataka, India

Academic Year 2019–2020

S.D.M. College of Engineering and Technology,
Dharwad–580002

CERTIFICATE

This is to certify that the project titled **Crop Price Prediction** is a bona fide work carried out by **Mahesh Mandakki (2SD16CS052), Vadiraj Raichur(2SD16CS131), Neha Torgal(2SD17CS407)** submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering** in **Computer Science and Engineering** of **S.D.M. College of Engineering and Technology, Dharwad, Karnataka** (An autonomous institution affiliated to Visvesvaraya Technological University, Belgaum, Karnataka), during the year 2019–2020. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The project has been approved, as it satisfies the academic requirements in respect of project report prescribed for the said degree.

Prof. Indira Umarji	**Dr. U. P. Kulkarni**
Project Guide	**Head of Department**

External Viva

Name of Examiners	Signature with Date
1)	
2)	

DECLARATION

We hereby declare that the dissertation work titled **CROP PRICE PREDICTION**, has been carried out under the guidance of **Prof. Indira Umarji, Assistant Professor, Department of Computer Science and Engineering, S.D.M. College of Engineering and Technology, Dharwad**, in partial fulfillment of the degree of **Bachelor of Engineering in Computer Science and Engineering** from **Visvesvaraya Technological University, Belgaum, Karnataka**, during the academic year 2019–2020.

I also declare that I have not submitted this dissertation to any other university for the award of any other degree.

Place: Dharwad

Date:

Name of Student	Signature with Date
1) Mahesh Mandakki (2SD16CS052)	
2) Vadiraj Raichur (2SD16CS131)	
3) Neha Torgal (2SD17CS407)	

ACKNOWLEDGEMENT

We consider it a privilege to express our sincere gratitude and respect to all those who guided and inspired us throughout this project. We express our heartfelt thanks to our guide **Prof. Indira Umarji**, for her valuable suggestions and guidance during the course of this project. The successful completion of this project owes to her coordination and streamlining of the project progress.

Abstract

Price Prediction, now a days has become very important agricultural problem which is to be solved only based on the available data. Data Mining techniques can be used to solve this problem. This work is based on finding suitable data models that helps in achieving high accuracy and generality for price prediction. For solving this problem, different Data Mining techniques were evaluated on different data sets.Agriculture mainly depends on him. Even then the farmers cannot predict prices for their commodities because prediction of prices plays a major challenge. Several characteristics are taken into account so that the crop price forecast is accurate. We consider the attributes of the Mysore region to make it a real-time application framework. Price prediction is a big issue for farmers who are not aware of the market prices. Forecasting price of agriculture commodities helps the agriculturist and also the agriculture department of mysore region to make decisions. The new model predicts the accuracy for the agricultural yields and it also avoids the role of middle man.

Problem Statement

From the perspective of agricultural business, the market price of certain crop reflects the demand of that crop in current stage. Therefore, to track and to forecast the market prices are both important tasks in agri-management, by which the production schedule can be adjusted to increase the profit by predicting the crop price.

Table of Contents

List of Figures

Chapter 1

Introduction

Predicting the price a given crop will yield in the future is extremely valuable when determining which types of crops to encourage . This static data is the crop production and data related to demands of various crops obtained from various government websites.The proposed system applies machine learning and prediction algorithm to identify the pattern among data and then process it as per input conditions. This in turn will propose the best feasible crops according to given environmental conditions

Some of the basic features of this project are as follows:

1. Enter the commodity name to get its price.
2. It will be showing future prediction also.
3. Graphical representation of all commodities.
4. Each algorithm analyzed root mean error difference is given.
5. Each commodity with all three (Linear,Ridge,Lasso) algorithms representation.

Country's Gross Development predominantly lean on Agriculture. Sixty percentage 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. Thus, heading towards the growth in Farmers' and Country's economy. 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 Forecasting System is to ensure that the Farmers make a better-informed decision and manage the price risk. 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.

Chapter 2

Design

2.1 System Architecture

System Architecture

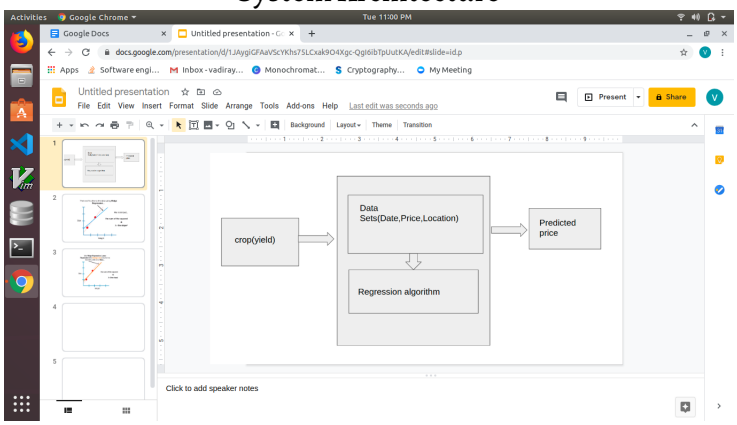

Figure 1: System Architecture

The System architecture of the crop price prediction is as show above in the figure i.e

1. The crop yields.
2. The yield crop are made available in datasets.
3. That datasets are undergoes regression algorithm.
4. To produce crop price prediction.

Chapter 3

Requirements

3.1 Project Specific Requirements

3.1.1 Functional Requirements

1. Easy registration and login.
2. Analyzed data representation.
3. User can view updated details.
4. Maintaining a database to record.

3.1.2 Non-Functional Requirements

1. The application must be reliable. As soon as the data is entered in the application, it must be immediately stored in the database.
2. The data entered, which is to be stored in the database must be stored accurately. The data when retrieved must be as same as the data stored.
3. The applications must be easy to access and easy to use.
4. The software must execute in any digital or a computing platform.

3.2 Software Requirements

1. MYSQL Database
2. Anaconda

3. Visual studio code editor
4. HTML,CSS,JavaScript,python
5. Web browser

3.3 Hardware Requirements

1. 2.2 GHz Processor
2. 4GB RAM
3. 16GB Hard Drive

Chapter 4

Implementation

Crop price predictionis a web based project.We have used JavaScript, HTML for front end and python for backend.We have also used CSS to make our page interactive and responsive.We have used MySQL database to store the data.As we are having datasets(available in kaggle.com) so we have displayed in our frontend.we are using three algorithm linear,ridge,lasso here.First implementing linear algorithm using datasets with its required formula.Displaying the output in a graphical representation and apply same for remaining two more algorithms and presenting it.The output of graph i.e, root mean square of each commodity with it's respective algorithm is displayed in command prompt.

4.1 Algorithm

Algorithms play important role in Machine learning or in data mining etc..It is having huge applications nowdays.Here it is about predicting the price.There are many algorithms to study but now we are using regression algorithm here.

4.1.1 Regression Algorihtm

Regression algorithms fall under the family of Supervised Machine Learning algorithms which is a subset of machine learning algorithms. One of the main features of supervised learning algorithms is that they model dependencies and relationships between the target output and input features to predict the value for new data. Regression algorithms predict the output values based on input features from the data fed in the system. The go-to methodology is the algorithm builds a model on the features of training data and using the model to predict value for new data. According to Oracle, here's a great defi-

nition of Regression – a data mining function to predict a number. Case in point, how regression models are leveraged to predict real estate value based on location, size and other factors. Today, regression models have many applications, particularly in financial forecasting, trend analysis, marketing, time series prediction and even drug response modeling. Some of the popular types of regression algorithms are linear regression, regression trees, lasso regression and multivariate regression

4.1.1.1 Linear Algorithm

It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear. Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line). It is represented by an equation Y=a+b*X + e, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

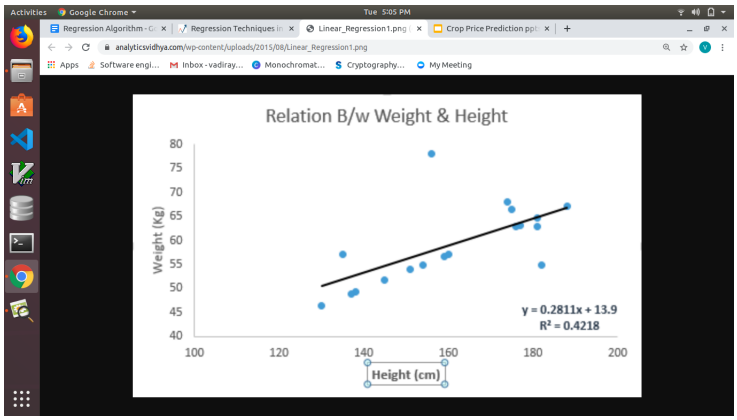

Figure 2: Linear Graph

The difference between simple linear regression and multiple linear regression is that, multiple linear regression has (>1) independent variables, whereas simple linear regression has only 1 independent variable.

4.1.1.2 Ridge Algorithm

Ridge Regression is a technique used when the data suffers from multicollinearity (independent variables are highly correlated). In multicollinearity, even though the least

squares estimates (OLS) are unbiased, their variances are large which deviates the observed value far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. Above, we saw the equation for linear regression. Remember? It can be represented as: y=a+ b*x This equation also has an error term. The complete equation becomes: y=a+b*x+e (error term), [error term is the value needed to correct for a prediction error between the observed and predicted value] => y=a+y= a+ b1x1+ b2x2+….+e, for multiple independent variables. In a linear equation, prediction errors can be decomposed into two sub components. First is due to the biased and second is due to the variance. Prediction error can occur due to any one of these two or both components. Here, we'll discuss about the error caused due to variance.

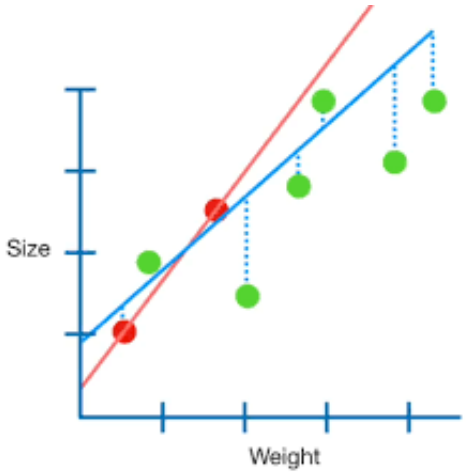

Figure 3: Sample Ridge

Ridge regression solves the multicollinearity problem through shrinkage parameter (lambda). Look at the equation below. In this equation, we have two components. First one is least square term and other one is lambda of the summation of 2 (beta- square) where is the coefficient. This is added to least square term in order to shrink the parameter to have a very low variance.

Important Points:

1. The assumptions of this regression is same as least squared regression except normality is not to be assumed.

2. Ridge regression shrinks the value of coefficients but doesn't reaches zero, which suggests no feature selection feature.

3. This is a regularization method and uses l2 regularization.

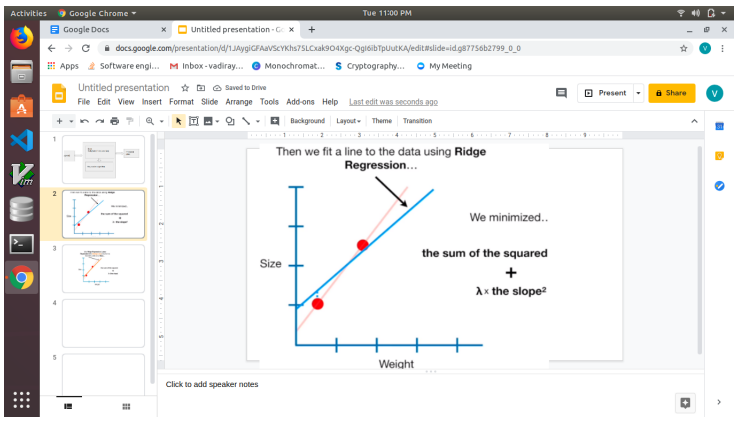

Figure 4: Ridge Graph

4.1.1.3 Lasso Algorithm

Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients. In addition, it is capable of reducing the variability and improving the accuracy of linear regression models. Look at the equation below: Lasso regression differs from ridge regression in a way that it uses absolute values in the penalty function, instead of squares. This leads to penalizing (or equivalently constraining the sum of the absolute values of the estimates) values which causes some of the parameter estimates to turn out exactly zero. Larger the penalty applied, further the estimates get shrunk towards absolute zero. This results to variable selection out of given n variables.

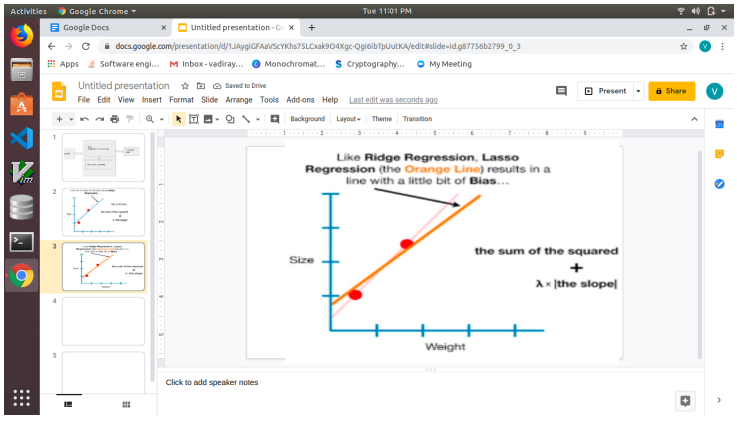

Figure 5: Lasso Graph

Important Points:

1. The assumptions of lasso regression is same as least squared regression except normality is not to be assumed.

2. Lasso Regression shrinks coefficients to zero (exactly zero), which certainly helps in feature selection.
3. Lasso is a regularization method and uses l1 regularization.
4. If group of predictors are highly correlated, lasso picks only one of them and shrinks the others to zero.

4.1.1.4 Difference of Algorithms

Algorithms		
Linear	Ridge	Lasso
It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear. Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).	Ridge Regression is a technique used when the data suffers from multicollinearity (independent variables are highly correlated). In multicollinearity, even though the least squares estimates (OLS) are unbiased, their variances are large which deviates the observed value far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. Above, we saw the equation for linear regression. Remember? It can be represented as: y=a+ b*x This equation also has an error term. The complete equation becomes: y=a+b*x+e (error term), [error term is the value needed to correct for a prediction error between the observed and predicted value] => y=a+y= a+ b1x1+ b2x2+. . . +e, for multiple independent variables.	Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients. In addition, it is capable of reducing the variability and improving the accuracy of linear regression models. Look at the equation below: LassoLasso regression differs from ridge regression in a way that it uses absolute values in the penalty function, instead of squares. This leads to penalizing (or equivalently constraining the sum of the absolute values of the estimates) values which causes some of the parameter estimates to turn out exactly zero. Larger the penalty applied, further the estimates get shrunk towards absolute zero. This results to variable selection out of given n variables.

4.2 Snapshots

This is the Login home page.

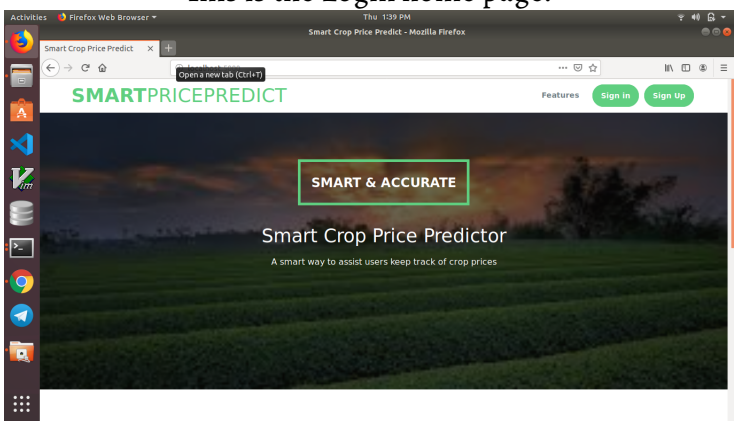

Figure 6: Register Page

Graphical representation of commodities.

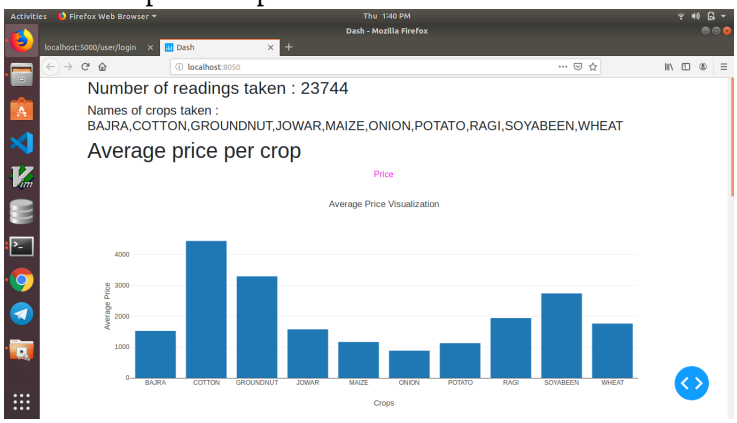

Figure 7: Graph Page 1

Graphical representation of count of commodities.

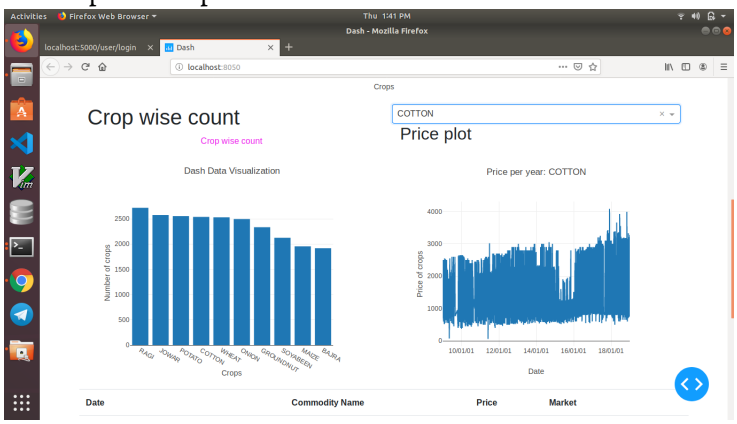

Figure 8: Graph Page 2

Output of Linear Regression Algorithm

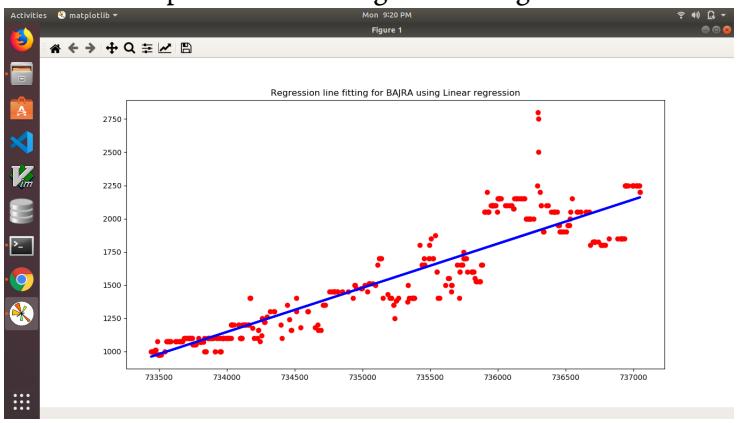

Figure 9: Linear Regression

Output of Ridge Regression Algorithm

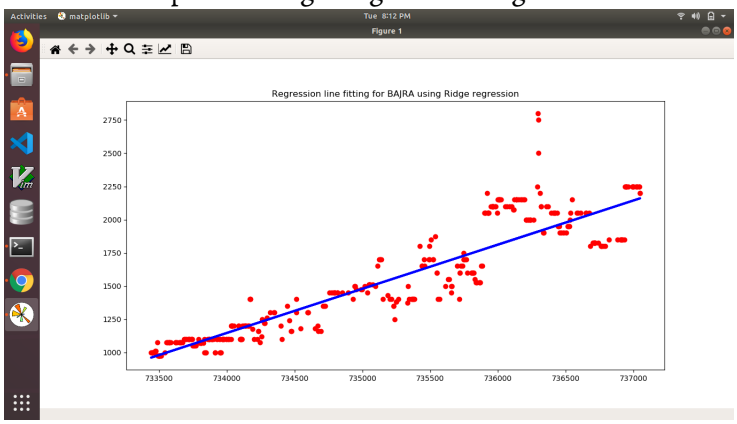

Figure 10: Ridge Regression

Output of Lasso Regression Algorithm

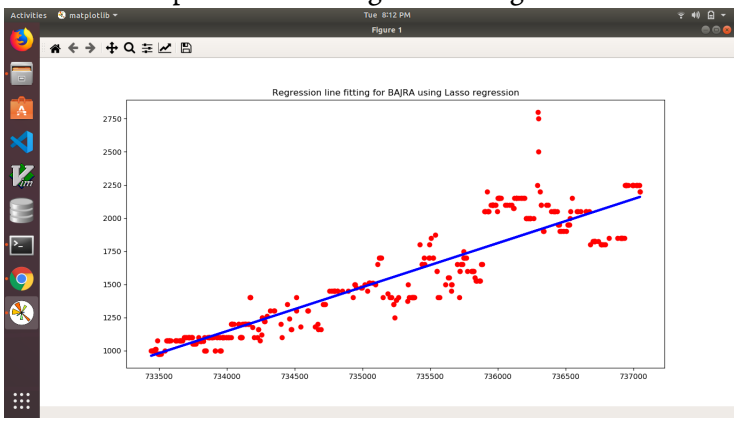

Figure 11: Lasso Regression

Output (Root mean sq error)

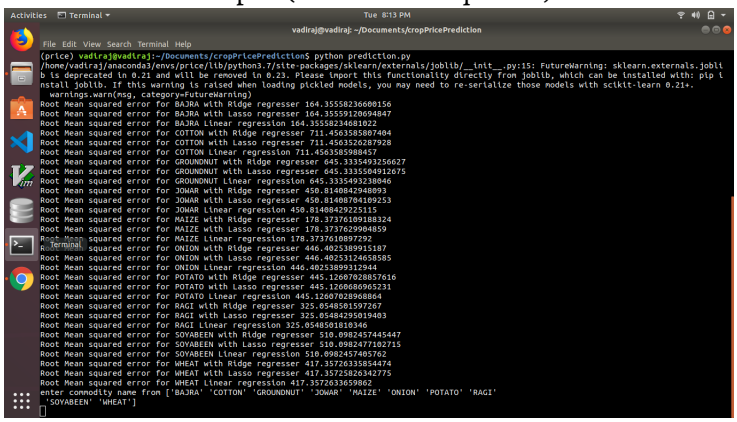

Figure 12: OUTPUT

Chapter 5

Test Cases

1. An alert message is displayed saying FAILURE! if invalid login Credentials are provided.
2. An alert message is displayed saying commodity name is not available here.
3. Have to run a correct name file if not Failure! Enter the right filename again command prompt it will ask.
4. If 'ctrl+c' is pressed while writing ruuning IT will end.
5. While showing of graph any wrong button is pressed all graph will stop displaying them.

Chapter 6

Conclusion

The research aims at predicting both the price and profit of the given crop before sowing. This web application runs on efficient machine learning algorithms and technologies having an overall user-friendly interface to the users. The training datasets so obtained provide the enough insights for predicting the appropriate price and demand in the markets. Thus, the system helps the farmers in reducing their difficulties and stop them by attempting suicides. . As the system lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.

Chapter 7

References

1. "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets", Yung-Hsing Peng, ChinShun Hsu, and Po-Chuang Huang, 2017 IEEE.

2. "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques", Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, 2015 IEEE.

3. Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining approach", 2015 International Conference on Computational Intelligence and Communication Networks.

4. R.Nagini, Dr. T.V. Rajnikanth, B.V. Kiranmayee, "Agriculture Yield Prediction Using Predictive Analytic Techniques , 2nd InternationalConference on Contemporary Computing and Informatics (ic3i),2016.

5. Machine Learning Algorithms 2nd Edition by Giuseppe Bonaccorso.