CROP PATTERN PREDICTION USING MACHINE LEARNING

MINOR PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OFTHE REQUIREMENTS FOR

THE AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

Information Technology

Submitted By:
Shubhamdeep Singh(1805556)
Sidhique Chaudhary(1805557)
Tanvi(1805566)

Submitted To:
Prof.Mohanjit Kaur Kang
Assistant Professor
Minor Project Coordinator

Department of Information Technology

Guru Nanak Dev Engineering College,

Ludhiana-141006

Abstract

The history of Agriculture in India dates back to Indus Valley civilization Era and even before that in some parts of Southern India.India ranks second worldwide in farm outputs. The share of agriculture in GDP increased drastically with about 41.49broadest economic sector and plays a significant role in the overall social-economic fabric of India. Today the farmers develop crops dependent on the experience picked up from the past age .Since the customary technique for cultivating is polished there exists an overabundance or shortage of yields without gathering the real necessity. The farmers don't know about the interest that happens in the current horticultural economy. This results in the misfortune to the Farmers. The communicated thought processes arranged by significance in the back of Farmer suicides have been condition, low produce costs, weight and hover of relative's obligations, poor water system, and blast inside the cost of development. The primary reason is the low costs of the items and the expanded expense of development. The expenses of yields are controlled by economic interest and the points of comment of the creation. Yield forecast is one of the undertakings that should be possible by bleeding edge ML calculations. The reason for this decline in the agriculture sector is due to the fact that farmers are not empowered and due to lack of application of IT in the farming sector. Farmers have less knowledge about the crops they grow. We tend to overcome this obstacle by applying machine learning techniques to predict the crop yield and name by considering various factors such as temperature, rainfall, season and area.

Acknowledgement

We would dearly thank our guide Prof. Mohanjit Kaur Kang who provided us extensive guidance and expertise throughout the study. We would also thank Dr K.S. Maan (H.O.D. of I.T. department) and our worthy principal Dr Sehajpal Singh for providing us this wonderful opportunity for discovering our potential.

Contents

Chapter 1

Introduction

1.1 Introduction to Project Page

Agriculture is the backbone of every economy. In a country like India, which has ever increasing demand of food due to rising population, advances in agriculture sector are required to meet the needs. Predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Implementation of such a system with an easy-to-use web based graphic user interface and the machine learning algorithm will be carried out. The results of the prediction will be made available to the farmer.Thus, for such kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield. Random forest algorithm is used. By analysing all these issues and problems like weather, temperature, humidity, rainfall, moisture, there is no proper solution and technologies to overcome the situation faced by us. Generally, data mining is the process of analysing data from various viewpoint and summarizing it into important information. Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees during training time and generating output of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

1.2 Project Category(Internet based, Application or System Development, Research based, Industry Automation, Network or System Administration)

It is an easy-to-use web based graphic user interface and the machine learning algorithm are carried out.

1.3 Objectives

The main objectives are:

- Use machine learning techniques to predict crop yield.
- Increase the accuracy of crop yield prediction.
- Analyse different climatic parameters (cloud cover, rainfall, temperature)

1.4 Problem Formulation

To Design, Develop and Implement the training model by using different inputs data. So machine will able to learn the features and extract the crop yield from the data by using machine leaning techniques.

1.5 Unique Features of the System

This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Implementation of such a system with an easy-to-use web based graphic user interface and the machine learning algorithm will be carried out.

Chapter 2

Requirement Analysis and System Administration

2.1 Requirements

This system architecture focuses 3 parts such as flow data, Machine learning techniques, and modules for detecting crop yield and feature selection modules. Polynomial regression In this model regression analysis is done by deriving the relationship between x independent variable and y dependent variable as nth degree polynomial of x. Decision Tree As decision tree employs greedy method, attribute chosen in the first step cannot be used later to give better classification of data. If at all it is used in the next steps Decision Tree over fits the training dataset that can lead to poor results Random Forests Random forest, as the name says it is a combination of number of decision trees and an ensemble classification model. Random forest model collects trained data from all the tree nodes and separates the weaker nodes training data to get better predictions. Both classification and regression problems are solved using RF model.

2.2 Expected hurdles

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 project more efficient to get the desired results.
2. Model must be trained by taking consideration of less computational efficiency and power.
3. Increase of error rate due to dynamically changing the environment.

Chapter 3

System Design

3.1 Design Approach (Function oriented or Object oriented)

The design approach is Object oriented. In design approach, the system state is decentralized among the objects and each object manages its own state information. For this we divided the project into various modules. There are various Modules based on the various parameters(object oriented). The following are the same:-

1. Data-sets: based on the data-sets taken for model preparations and preprocessing. It is divided into two parts –Weather Data: District wise Data of each state of Punjab and others csv files for joining and preprocessing.
2. Encoded Files: Based on encoding of the data of crops, district and seasons.
3. Images: Results of the various models and plots.
4. Saved Modules: Saved Models used for testing the data.
5. text files: Basis details and requirements of the project.
6. Last section are all the python files used in the project.

3.2 Detail Design

The system design consists of raw data collected from Nasas site. Here the data of each state of Punjab in the datasets. Then the after preprocessing of the data is joined in one file and then they are encoded for the model preparation. The encoded data is taken as the data and the whole dataset is splitted into two parts of training and testing data. The training data is used as input for various models (KNN, Random Forest, Grad Boost, and Linear Regression)

and the models undergo learning phase. After this the data is tested against the testing data. The predicted value and the actual data is compared. For all the models accuracy and other errors are compared. The model with highest accuracy and least error is taken for testing the cropping pattern.

3.3 SystemDesign using various structured analysis and design tools such as : DFS's, Data Dictionary, Structured charts, Flowcharts or UML

STRUCTURED CHART:

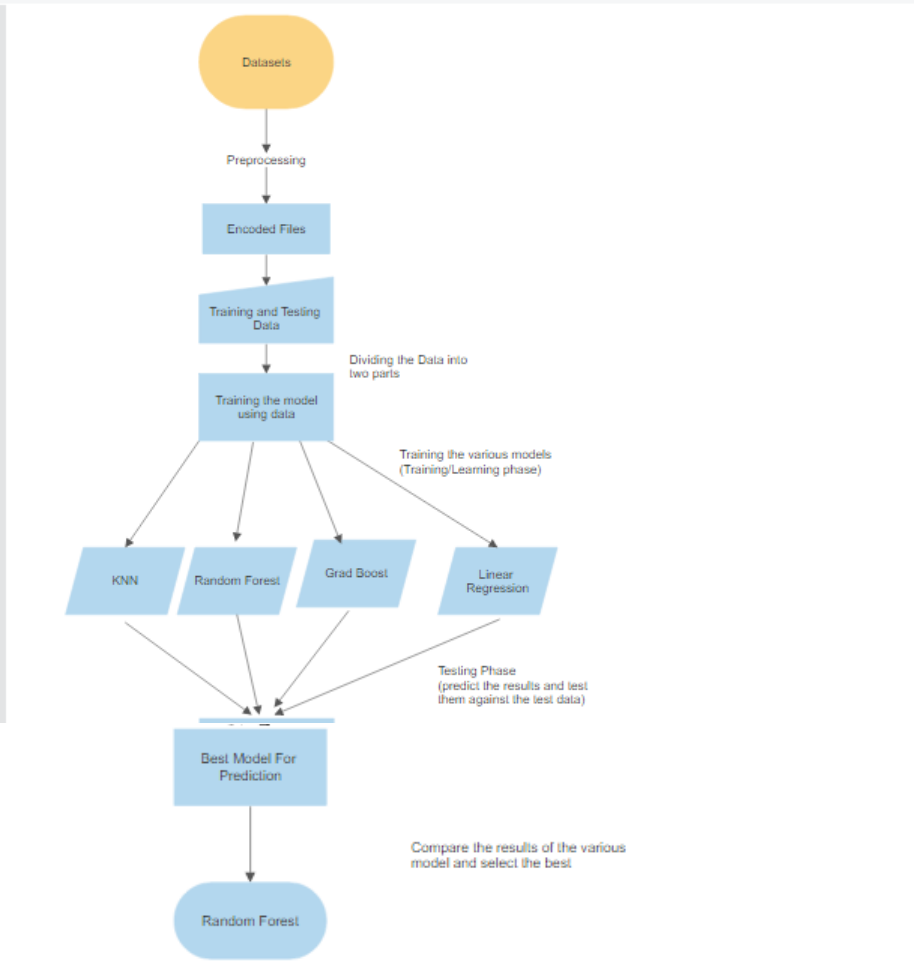

Figure 1: Flow Chart of the system Design used for the project

3.4 User Interface Design

User Interface used for the project is pycharm in Darcula theme. It is indirectly an Graphical user Interface. All the results of the user interface are stored in the Saved Models (for Models) result in the form of images are stored in Images Folder and result of the csv file format is stored in various Folder as per there use. As such there is no use of user interface for output section, it is only for the input section.

Chapter 4

Implementation, Testing and Maintenance

4.1 Introduction to Languages, IDE's, Tools and Technologies used for Implementation

This model gives clear picture of huge amount of data capture and preprocessing of data to remove the unwanted data such as NULL etc presented in it. During preprocessing step we spit the dataset into training and testing dataset. Train dataset to detect the crop yield present in the dataset using appropriate supervised learning algorithms. Apply the machine learning techniques which are helpful for finding crop yield for any of new data occurred in the data. After this data acquisition suitable machine learning algorithm must be applied to compute efficiency and capability of the model, here we have applied various machine learning algorithms like random forest, Polynomial Regression, Decision Tree etc. Metrics like accuracy, precision will be calculated for the proposed model. This system architecture focuses 3 parts such as flow data, Machine learning techniques, and modules for detecting crop yield and feature

4.2 Testing Techniques and Test Plans

There will be 70 percentage of data for training and 30 percent for testing. Initially we are choosing samples for training of model and chosen some of the features like Yield, temperature and Location etc. After successful training and the testing of data-set we moved further for finding the accuracy the model. The accuracy of the model shows that how we have predicted the yield of the crop in compare to original data the more the accuracy of the model is near to the original yield value. Data analysis is done by plotting the yield variable with different attributes like state, pressure, temperature Max, temperature Min etc.

Chapter 5

Results and Discussions

Machine learning frameworks played an important role in performing this experiment. The Results shows that we can attain an accurate crop yield prediction using the Random Forest algorithm. Random Forest algorithm achieves a largest number of crop yield models with a lowest models.It is suitable for massive crop yield prediction in agricultural planning. This makes the farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas.

5.1 User Interface Representation (of Respective Project)

User Interface used for the project is pycharm in Darcula theme.It is indirectly an Graphical user Interface. All the results of the user interface are stored in the Saved Models (for Models) result in the form of images are stored in Images Folder and result of the csv file format is stored in various Folder as per there use. As such there is no use of user interface for output section, it is only for the input section.

5.1.1 Brief Description of Various Modules of the system

There are various Modules based on the various parameters. The following are the same:

1. Datasets: based on the datasets taken for model preparations and preprocessing. It is divided into two parts –Weather Data: District wise Data of each state of Punjab and others csv files for joining and preprocessing.

2. Encoded Files: Based on encoding of the data of crops, district and seasons.

3. Images: Results of the various models and plots.

4. Saved Modules: Saved Models used for testing the data.
5. Text files: Basis details and requirements of the project.
6. Last section are all the python files used in the project.

5.2 Snapshots of system with brief detail of each

5.3 Back Ends Representation (Database to be used)

No Database in used for this Machine Learning project as backend.

5.3.1 Snapshots of Database Tables with brief description

No Database used in this project.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

This project is undertaken using machine learning and evaluates the performance by using Random forest, Polynomial Regression and Decision Tree algorithms. In our proposed model among all the three algorithm Random forest gives the better yield prediction as compared to other algorithms. Along with random forest, Polynomial Regression, Decision Tree model classify the output that shows improvements in dataset. So we analyzed that proposed model has got more efficiency than the existing model for finding crop yield.

6.2 Future Scope

The scope of the project is to determine the crop yield of an area by considering dataset with some features which are important or related to crop production such as temperature, moisture, rainfall, and production of the crop in previous years. To predict a continuous value, regression models are used. It is a supervised technique. The coefficients are preprocessed and fit into the trained data during training and construction the regression model. The main focus here is to reduce the cost function by finding the best fit-line. The output function facilitates in error measurement.

References

[1] http://www.ijstr.org/final-print/jan2020/Design-And-Implementation-Of-Crop-Yield-Prediction-Model-In-Agriculture.pdf

[2] D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015. 7.

[3] Yethiraj N G , " Applying data mining techniques in the field of Agriculture and allied sciences", Vol 01, Issue 02, December 2012