CROP YIELD PREDICTION USING MACHINE LEARNING

A PROJECT PHASE 1

submitted by

PRASATH.E (19205034)

ARUN.A (19205002)

RAMESH.S (19205037)

VIGNESH.R (19205052)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

ERODE SENGUNTHAR ENGINEERING COLLEGE

(An Autonomous Institution)

PERUNDURAI, ERODE – 57.

ANNA UNIVERSITY :: CHENNAI 600 025

November 2022

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "**CROP YIELD PREDICTION USING MACHINE LEARNING** " is the bonafide work of "**PRASATH E (19205034) ,ARUN A (19205002), RAMESH S (19205037) , VIGNESH R (19205052) "** who carried out the project work under my supervision.

SIGNATURE **SIGNATURE**

HEAD OF DEPARTMENT
Department of Information Technology,
Erode sengunthar engineering college
Erode-638 057.

HEAD OFDEPARTMENT
Department of Information Technology,
Erode sengunthur Engineering College,
Erode-638 057.

Submitted for the Anna University Viva Voce examination held on_____ for **CS8611 – Project Work** during the academic year 2021-2022.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

Abstract

India is generally an agricultural country. Agriculture is the single most important contributor to the Indian economy. Agriculture crop production depends on the season, biological, and economic cause. The prognosticating of agricultural yield is challenging and desirable task for every nation. The Farmers are struggling to produce the yield because of unpredictable climatic changes and drastically reduce in water resource so; we are creating an agriculture data. This data could be gathered, stored and analyzed for useful information. It is a significant field for determining and examining the yield. The indispensable part of the cultivator is to think about the creation of the yield. Long previously, estimating was completed by considering the cultivator's past experience on the chose zone. The estimating was the significant measures which should be addressed by thinking about the information available. This examination assists with suggesting a model for anticipating the yield from the previous information. At present the forecast of the Crop assists with taking care of the farming issue. All the ranchers by and large needs to know the desire for the yield dependent on his experience By utilizing Data mining technique, the improved determination should be possible. Different Data Mining strategies have been utilized for figuring the impending year's creation. It is used to promote new advanced methods and approaches such as data mining that can give the information of the previous results to the crop yield estimation.

ACKNOWLEDGEMENT

First of all, we thank the almighty God for the blessings and support given to us to complete this project successfully.

We would like to express our deep thanks to our beloved president & correspondent **Mr.G.KAMALAMURUGAN.,** for giving motivation and providing all necessary facilities for successful completion of the project.

We take this opportunity to express our profound gratitude and deep regards to **Mr.S.N.THANGARAJU,** Secretary, for his constant encouragement throughout the course of our work.

With pleasure, we would like to express our gratefulness to **Dr.V.VENKATACHALAM,B.E.,M.S.,M.Tech.,Ph.D.,** Principal, for the moral support and deeds in bringing out this project successfully.

We express our gratitude to **Dr.**Head of Department, Department of Information Technology for encouraging a lot throughout this work.

We hearty extend our thanks to our project coordinator **Dr.** and our guide **Dr.** for their support and guidance throughout the project for the successful completion and enhancement of the work.

We express our thanks to all our Teaching and Non-Teaching staff members for their advice and encouragement to do the project work with full interest and enthusiasm.

Finally, we would like to thank all our friends and family members for providing us a constant support to complete this project work.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

1.1 LINEAR REGRESSION:

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.

1.2 OBJECTIVE:

We are in an age often referred to as the information age. In **t**his information age, because we believe that information leads to power and success, and thanks to sophisticated technologies such as computers, satellites, etc., we have been collecting tremendous amounts of information.Initially, with the advent of computers and means for mass digital storage, we started collecting and storing all sorts of data, counting on the power of computers to help sort through this amalgam of information.Data Mining is extraction of unknown information from huge data It is a powerful new technology with great potential to help companies focus on the most important information. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions.In this project, we are going to discuss yield loss mechanisms, yield analysis and common physical design methods to improve yield. Yield is defined as the ratio of the number of products that can be sold to the number of products that can be manufactured.

CHAPTER 2

LITERATURE SURVEY

Literature survey:1

ML Methods for Crop Yield Prediction and Estimation: An Exploration

DESCRIPTION:

Machine learning Has performed a essential position within the estimation of crop yield for both farmers and consumers of the products. Machine learning techniques learn from data set related to the environment on which the estimations and estimation are to be made and the outcome of the learning process are used by farmers for corrective measures for yield optimization. This paper we explore various ML techniques utilized in crop yield estimation and provide the detailed analysis of accuracy of the techniques.

KEYWORDS: Machine Learning, Crop Yield

LIMITATIONS:

In this paper we implements ANN, SVM and decision forest technique on our dataset to determine the accuracy of the techniques. The techniques were used to predict the yield production of rice, potatoes' and wheat in Chennai of Tamil Nadu. The parameters considered are moderate rainfall, humidity, urea, temperature and soil.

BENEFITS:

To high and less warmth, common drizzle, dankness, clime kinds of area sorts of alchemical fertilizer, sorts of clay, sand structure, clay balance, clay moisture, clay rational, clay response and clay character.

CONCLUSION:

In this paper, we have explored some techniques used in the estimation of crop yield for promoting food security and economic growth of individual country. Some techniques like decision forest, SVM and ANN were evaluated using some dataset based on some parameters and the accuracy of the methods was presented under the result section of this paper.

Literature Survey:2

A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction

DESCRIPTION:

An early and reliable estimation of crop yield is essential in quantitative and financial evaluation at the field level for determining strategic plans in agricultural commodities for import-export policies and doubling farmer's incomes. Crop yield predictions are carried out to estimate higher crop yield through the use of machine learning algorithms which are one of the challenging issues in the agricultural sector. Due to this developing significance of crop yield prediction, this article provides an exhaustive review on the use of machine learning algorithms to predict crop yield with special emphasis on palm oil yield prediction. Initially, the current status of palm oil yield around the world is presented, along with a brief discussion on the overview of widely used features and prediction algorithms. Then, the critical evaluation of the state-ofthe-art machine learning-based crop yield prediction, machine learning application in the palm oil industry and comparative analysis of related studies are presented. Consequently, a detailed study of the advantages and difficulties related to machine learning-based crop yield prediction and proper identification of current and future challenges to the agricultural industry is presented. The potential solutions are additionally prescribed in order to alleviate existing problems in crop yield prediction. Since one of the major objectives of this study is to explore the future perspectives of machine learning-based palm oil yield prediction, the areas including application of remote sensing, plant's growth and disease recognition, mapping and tree counting, optimum features and algorithms have been broadly discussed. Finally, a prospective architecture of machine learning-based palm oil yield prediction has been proposed based on the critical evaluation of existing related studies. This technology will fulfill its promise by performing new research challenges in the analysis of crop yield prediction and the development of an extremely effective model for the prediction of palm oil yields with the most minimal computational difficulty.

KEYWORDS: Artificial intelligence, crop yield prediction, deep learning, machine learning, palm oil yield.

LIMITATIONS:

Among these sub-features, precipitation, vapor pressure deficit, humidity, solar radiation, temperature, rainfall, and vapor pressure have been widely employed by researchers.

BENEFITS:

The rapid expansion of oil palm plantations has also led to deforestation and a series of negative environmental impacts, such as forest estate losses, social costs, alternative revenue losses, reduced biodiversity, and diminished ecological connectivity.

CONCLUSION:

In order to feed a rising world population, new technology in the agricultural industry needs to be implemented. Apart from this, agriculturists need a proper guideline in time that will allow them to forecast crop yields so that they can formulate effective strategies to maximize crop yields. ML frameworks offer a clear insight into the process by assessing the vast sets of data and interpreting the obtained information. The models describing the correlations between constituents and actions are built through these technologies. In addition, the future reactions in a given situation can also be predicted through the ML models. The present review illustrates that a wide range of attributes is utilized by the selected articles, focusing on the data availability and research scope. Most of the referred articles explore yield forecasting through the ML algorithms.

Literature survey : 3

Applying Big Data for Intelligent Agriculture-Based Crop Selection Analysis

DESCRIPTION:

The growth of the human population comes the constantly rising demand for agricultural products. Nevertheless, as the world experiences climate change, many crops are often damaged by weather conditions.This study utilizes Intelligent Agriculture IoT equipment to monitor the environmental factors on a farm. The collected data underwent 3D cluster analysis to yield analysis of the environmental factors of that farm. The proposed scheme bears the following features: (1) data normalization is achieved via the combination of moving average and average variance; (2) we applied 3D cluster analysis to analyze the relation between environmental factors and subsequently examine the rules of thumb held by the farmers; (3) the system determines whether a selected crop has been placed in the appropriate cluster; and (4) the system sets a critical value in the cluster based on future environments and provides advice on whether a crop is suitable for the farm. We placed Intelligent Agriculture IoT equipment in the farm for monitoring purposes and ran an actual-scenario analysis using the algorithm in our study; results confirm that our proposed scheme is indeed feasible.

KEYWORDS: Big Data, Intelligent Agriculture, Internet of Things, Agricultural engineering, Data mining.

LIMITATION:

It utilizes cloud technology to store IoT sensor data, which are then used in big data analysis for farm management.

BENEFITS:

This elaborates on our Intelligent Agriculture experiment results and big data experiment results . This work was financially supported from the Young Scholar Fellowship Program by Ministry of Science and Technology.

CONCLUSION:

Faced with extreme climate changes and increased global population, we are forced to emphasizemust address food issues including such as crops and agriculture. Our study proposes using an Intelligent Agriculture platform to monitor the environmental factors at on a farm and applying these environmental factors in analyzing cultivation techniques held by farmers. Our proposed scheme employs moving average and variance in data cleaning, which cleans out data with more drastic variation. Our study applies autocorrelation to compute periodicity while using 3D cluster correlation to conduct behavior analysis of farmer actions such as application of fertilizer or pesticide. Our study takes looks into environmental factors to assess whether a crop is suitable for a farm; it also takes global warming into consideration. The eExperiment results demonstrate show our analysis of four crops using our proposed approach; the results prove that farmers can gain a better understanding of whether a crop is suitable for their farm through by looking into factors such as temperature and soil moisture content. TThrough the environmental factor analysis proposed in our study help, farmers can gain insight into which crops they can grow; meanwhile, while the system keeps track of and analyzes crop cultivation behavior. In the future, our proposed scheme can incorporate artificial intelligence and apply the analysis results to help farmers achieve automatic cultivation and environment control.

Literature survey : 4

Crop Yield Prediction Using Linear Support Vector Machine

DESCRIPTION:

He main objective of this proposal is to build a Machine Learning model that can accurately predict the rice crop yield prediction. Over 97% of the population in India depends on rice for food and is the second-highest in overall agriculture productions. But during recent years the farmers had suffered a huge loss in productions due to unexpected weather change, no knowledge about soil, underground water, area supported crops. As crop production depends on a lot of these factors, it is important to follow these factors for successful crop yield. So we are proposing a model that can accurately predict the crop yield. The algorithms used were the Support Vector Machine (SVM). SVM is used to classify the crop based on the factors of the area, season. And we are also implementing a Web Application that enables the users to interact with the ML model and make their prediction with their given inputs. The proposed system uses the Weka tool for creating the machine learning algorithms and Html, CSS, JavaScript for developing the web application.

KEYWORDS: Support Vector Machine (SVM), Weka, Rice crop yield prediction, Web application.

LIMITATION:

The model will mainly focus on crop production based on four factors and one Machine Learning algorithm called SVM (support vector machine). SVM is used to classify whether rice can grow in that area based on the data from soil, temperature, underground water and rainfall.

BENEFITS:

The only season recommended is Kharif. Our Machine Learning model which is trained with Support Vector Machine has managed to obtain an accuracy percentage of 96.5%.

CONCLUSION:

There is so much to explore in machine learning yet, as there can be new algorithms, new techniques in the future. Our paper is a simple crop prediction recommendation system. which is only limited to one state Tamil Nadu, as we hope to do more papers on other Indian states and encourage other fellow researchers to also persue research in the agriculture field, as this is our main source of food all over India. It alone contributes 60% of the entire GDP. But since 2018 it is gradually decreasing, the per capita water availability is also decreasing, which will result in a lot of crop production failures. And also there are multiple numbers of suicides of farmers all over India, who just work very hard and don't get the expected results due to many factors. This paper is a small contribution to the agriculture field and dedicated to all the farmers, to help them in their farming, so that they can get profits and benefits of the new technologies which they don't have any idea of. So finally we want to conclude that as an Engineer we should take responsibility and contribute our knowledge to the betterment of our society or country.

Literature survey : 5

Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis

DESCRITION:

Soil is an important parameter affecting crop yield prediction. Analysis of soil nutrients can aid farmers and soil analysts to get higher yield of the crops by making prior arrangements. In this paper, various machine learning techniques have been implemented in order to predict Mustard Crop yield in advance from soil analysis. Data for the experimental set-up has been collected from Department of Agriculture Department, Talab Tillo, Jammu; comprising soil samples of different districts of Jammu region for Mustard crop. For the current study, five supervised machine learning techniques namely K-Nearest Neighbor (KNN), Naïve Bayes, Multinomial Logistic Regression, Artificial Neural Network (ANN) and Random Forest have been applied on the collected data. To assess the performance of each technique under study; five parameters namely accuracy, recall, precision, specificity and f-score have been evaluated. Experimentation has been carried out to make known the most accurate technique for mustard crop yield prediction. From experimental results, it has been predicted that KNN and ANN (among the undertaken ML techniques for the study) found to be most accurate techniques for mustard crop yield prediction.

KEYWORDS: ANN, KNN, Machine Learning, Mustard Crop, Naïve Bayes, Random Forest.

LIMITATION:

Random Forest is a supervised learning algorithm. It creates a forest and makes it somehow random. Random Forest is a flexible, easy to use machine learning algorithm that produces, even**.**

BENEFITS:

It can be found out that all ML techniques under study can be used for crop yield prediction. KNN and random forest predicted highest accuracy of 88.67% and 94.13% respectively whereas Naïve Bayes predicted lowest accuracy of 72.33%. In terms of precision, ANN predicted highest value of 99.94% whereas Logistic regression predicted lowest value of 24.17%.

CONCLUSION:

The experimental study, it can be concluded that ML techniques can be effectively used for yield prediction of mustard crop. But, in this study, KNN and ANN are found to be most accurate techniques for mustard crop yield prediction. These effective ML techniques will help the farmers in predicting yield in advance based on soil parameters. In future, Crop yield prediction with huge soil data set can be implemented in Big Data environment. On the basis of results of yield prediction, fertilizer recommendations can also be implemented to help the soil analysts and farmers to take decisions accordingly in case of low crop yield prediction.

Literature survey : 6

ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUE TO PREDICT ANNUAL YIELD OF MAJOR CROPS

DESCRIPTION:

India is generally an agricultural country. Agriculture is the single most important provider to the Indian economy. Agriculture crop production depends on the season, organic, and monetary cause. The prognostication of agricultural yield is challenging and pleasing task for every nation. Nowadays, Farmers are hostile to produce the yield because of erratic climatic changes and scarcity of water resource. The main objective is collecting agricultural data which can be stored and analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, advanced methods can be introduced to predict crop yield and it also helps the farmer to choose the most suitable crop, thereby improving the value and gain of the farming area.

KEYWORDS: Data Mining, Classification, Crop Yield, Accuracy, K-Nearest Neighbor (KNN), Linear Regression.

LIMITATIONS:

The final results, means category effectiveness could be superior through the utilization of LR-PROPOSED being an optimization technique within the category method.

BENEFITS:

The efficient classification Linear Regression algorithm is used to develop the model. This algorithm is compared, and accuracy is evaluated.it is observed that Linear Regression had the best predictive power with high accuracy as compared to KNearest Neighbor.

CONCLUSION:

In accurate prediction of different specified crop yields across different districts will help to farmers of India. Yield estimation models are utilized in preciseness Agriculture to extend yield production to satisfy demand and to recommend to the government in regard to prediction crop yield on imports of Trichy, Tamilnadu dataset to avoid overlapping. During this work the regression approach were tested in their yield prediction capabilities. The readings were used for model inputs. Linear regression algorithms offered acceptable estimation accuracy, though higher prognostic power could also be obtained by parameters like year, crop, area, production (in tons) and alternative variables, like climate, agricultural practices and soil characteristics are including within the model development. The model using linear regression can be suggested for Ecuadorian conditions. In yield prognostic models are not existent for any crop. From this proposed system the yield of crop (sugarcane, cotton, and turmeric) are predicted in highest level. This model may be reformulated using alternative crop assessments within the future, to develop methods for increasing yield and land territorial management in alternative crops of importance, like wheat, rice

LITERATURE SURVEY:7

Crop Yield Prediction and Efficient use of Fertilizers

DESCRIPTION:

India being an agriculture country, its economy predominantly depends on agriculture yield growth and agroindustry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.

KEYWORDS: Artificial neural network, Random forest algorithm, Backpropagation algorithm, Prediction.

LIMITATIONS:

Prediction of the crop yield using the efficient algorithm and suggest how much quantity of fertilizer should be used to get the proper yield for the crop.

BENEFITS:

An effort is made in order to know the crop production analysis and is processed by implementing both the Random Forest algorithm and Backpropagation algorithm.

CONCLUSION:

Crop yield prediction and efficient use of the fertilizer is successfully predicted and also found the efficient algorithm from both the algorithm and obtained the most efficient output of the yield. In future developing the web application based on this ideology and make the user use this easily and help the user to understand the yield of the crop, he is going to crop in that season.

LITERATURE SURVEY:8

A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

DESCRIPTION:

Agriculture is defined as the science and art of cultivating the flora and fauna. Farming in India is ranked as second around the globe and occupies 60.45% of Indian land. The Indian economy, dominatingly, depends upon farming along with agro-industry things. The soil ingredients (like Nitrogen, Phosphorous, Potassium), crop rotation, soil clamminess, atmospheric and surface temperature, precipitation, etc, play an efficient role in cultivation. The present evidence related to this field includes a model which is incorporated with ML algorithms (Random Forest, Decision Tree, Artificial Neural Network) to determine best crop. In this paper, the proposed model is enhanced by applying deep learning techniques and along with the prediction of crop, a clear information is achieved regarding the amounts of soil ingredients needed with their expenses separately. It provides a better accuracy than the existing model. It analyzes the given data and help the farmers in predicting a crop which in return help in gaining profits. The climatic and soil conditions of land are taken into consideration to predict a proper yield. The objective is to present a python based system that uses strategies smartly to anticipate the most productive reap in given conditions with less expenses. In this paper, SVM is executed as Machine Learning algorithm while LSTM and RNN are used as Deep Learning algorithms.

KEYWORDS: Agriculture, Crop Prediction, Machine Learning, Deep Learning, SVM, LSTM, RNN

LIMITATIONS:

The machine learning and deep learning techniques are executed in order to predict the best crop production.

BENEFITS:

A model is constructed in which required AI algorithms are utilized which in return will provide the best suitable crop that should be grown on a particular land.

CONCLUSION:

The proposed model is constructed by using AI algorithms to reduce the farmers' problems of getting losses in their farms due to lack of knowledge of cultivation in different soil and weather conditions. The model is created by using machine learning (SVM) and deep learning (LSTM, RNN) techniques. The model predicts best crops that should be grown on land with less expenses among a number of crops available after analyzing the prediction parameters. To the best of studies, there is no such work in existence that uses the same techniques in predicting the crops. Hence, it is concluded that there is an enhancement in the accuracy of this research work when compared to the existing work that used another techniques for prediction of crops. The accuracy is calculated as 97%. It has a vast extension in future and can be actualized and interfaced with a flexible and multi-skilled application. The farmers need to be educated and hence, will get a clear information regarding best crop yield on their mobiles. With this, even if the rancher is at home, the work can be managed at that particular instant of time, without facing any kind of loss ahead. The progress in the agribusiness field will be extremely appreciable which will further result in helping the farmers in production of crops.

LITERATURE SURVEY: 9

PREDICTION OF CROP GROWTH USING MACHINE LEARNING BASED ON SEED FEATURES

DESCRIPTION:

The presence of plant species in the wrong place and time is identified as plant weeds. The loss of yield could result from interference with plant crop weed species. To classify the plant weeds among the seeds, seed classification is carried out in this paper. Here the image of seeds and datasets of sample seed are input. In pre-processing stage the seed image is given as input. An unwanted seeds are removed by comparing the features of seed with sample seeds features by using ID3 algorithm. One of the reasons for failure in crop yield production is selecting suitable soil for crop. As sample dataset contains the detail of growth of crop in soil, it will help for selecting the suitable soil for seeds. The features of sample dataset are compared with the features extracted from the affected crop and predict the disease and prevention measures taken place. In this method prediction is done only after the growth of crop which leads to decrease in quality of crop growth. In order to overcome these issues, diseases can be predicted using seed features by comparing the features with sample dataset. Using Support Vector Machine algorithm the seeds are classified based on the growth and predicting the diseases of crop. This is done by training the dataset by comparing the features extracted from new seeds and features of sample seeds and predicting the crop growth and diseases. Based on prediction of crop growth and crop diseases, preventive measures takes place.

KEYWORDS: Data Mining, Big Data, Feature Extraction, Crop Growth

LIMITATIONS:

Image processing is used to denoise the image and to extract the image characteristics. The ID3 algorithm collects the characteristics extracted from the sample seed data and locates the unwanted seeds.

BENEFITS:

The SVM performance with various filters, in both experiments. , in terms of PSNR, RMSE and MSE. he proposed method achieves higher percentage of accuracy with average filter than median and wiener filter.

CONCLUSION:

In this paper, seed classification is carried out in order to classify the plant weeds among the seeds. Here the seed image and the sample seed datasets are entered. The seed image is given as an input in the preprocessing stage. An undesirable seed is eliminated by comparing the seed characteristics with the samples of seeds with the ID3 algorithm. The selection of appropriate cultivated soil is one of the reasons why crop production fails. As the data set of samples contains information about crop growth in soil, it helps to select the appropriate soil for seeds.

LITERATURE SURVEY:10

County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model

DESCRIPTION:

Yield prediction is of great significance for yield mapping, crop market planning, crop insurance, and harvest management. Remote sensing is becoming increasingly important in crop yield prediction. Based on remote sensing data, great progress has been made in this field by using machine learning, especially the Deep Learning (DL) method, including Convolutional Neural Network (CNN) or Long Short-Term Memory (LSTM). Recent experiments in this area suggested that CNN can explore more spatial features and LSTM has the ability to reveal phenological characteristics, which both play an important role in crop yield prediction. However, very few experiments combining these two models for crop yield prediction have been reported. In this paper, we propose a deep CNN-LSTM model for both end-of-season and in-season soybean yield prediction in CONUS at the county-level. The model was trained by crop growth variables and environment variables, which include weather data, MODIS Land Surface Temperature (LST) data, and MODIS Surface Reflectance (SR) data; historical soybean yield data were employed as labels. Based on the Google Earth Engine (GEE), all these training data were combined and transformed into histogram-based tensors for deep learning. The results of the experiment indicate that the prediction performance of the proposed CNN-LSTM model can outperform the pure CNN or LSTM model in both end-of-season and in-season. The proposed method shows great potential in improving the accuracy of yield prediction for other crops like corn, wheat, and potatoes at fine scales in the future.

KEYWORDS: soybean; yield prediction; county-level; Google Earth Engine; CNN-LSTM

LIMITATIONS:

These factors usually have a complicated interaction with each other during the growing season. However, the impact of all these factors can be demonstrated by crop growing status finally; in other words.

BENEFITS:

Over time, more information was integrated into training data, and the model performance was improved gradually. Note that the RMSE of 2012 is usually relatively high than other years.

CONCLUSION:

In this paper, Accurate early yield prediction is of great significance for crop market planning, crop insurance, and harvest management. In this paper, a GEE-based CNN-LSTM model was proposed for both in-season and end-of-season soybean yield prediction by county-level in CONUS. From 2011 to 2015, the results demonstrate for the first time evidence that (1) compared with the CNN or LSTM, the prediction performance of the proposed CNN-LSTM model was proven to be the best. Based on the proposed method, the end-of-season yield prediction can obtain high accuracy with RMSE = 329.53 averaged from 2011 to 2015 and R 2 = 0.78 for five years together. (2) An early prediction on AUG 21st can achieve a satisfying result with RMSE = 353.74 and R 2 = 0.74, which is comparable to end-of-season result but can win a long time before USDA issue data. (3) The method is highly efficient, as it can benefit from the great computing power of GEE and a dimension reduction method. (4) MODIS surface reflectance played a more important role in the method than environmental feature

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

In review the study on the purpose of data mining methods in the area of agriculture. A few of the data mining methods, such as the svm, k nearest neighbor, support vector machines, artificial neural networks and applied in the area of agriculture were presented. An exact estimate of crop range and risk assists these business groups in planning, supply chain decision similar to production scheduling.It described they expect crop yields using data mining techniques.

DEMERITS:

- It does not execute very well when the data set has more sound i.e. target classes are overlapping.
- The number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.
- Accuracy depends on the quality of the data.
- With large data, the prediction stage might be slow.
- Sensitive to the scale of the data and irrelevant features.
- Require high memory – need to store all of the training data.
- Given that it stores all of the training, it can be computationally expensive.

3.2 PROPOSED SYSTEM:

The proposed work is to focus average Modified Linear Regression data from the 15 year period were plotted against the date of image acquisition and a quadratic model was fitted to visualize the progression of sugarcane crop growth. This model identified when maximum vigour of the annual crop was achieved, via the peak of the quadratic curve and, therefore, indicates when images should be captured to get

maximum vigour of crop and ultimately predict yield in the season.The vertex form of the quadratic model as shown in Equation was used to shift the vertical axis of the curve according to the acquired MLR value in maximum vigour period of a specific year. The KNN models have been experimented using different partitions of training patterns and different combinations of KNN parameters.

MERITS:

- Quick calculation time.
- Simple algorithm to interpret.
- Versatile useful for regression and classification.
- High accuracy you do not need to compare with better-supervised learning models.
- The main advantage of KNN over other algorithms is that KNN can be used for multiclass classification.