ANALYSIS OF CROP YIELD PREDICTION USING DATA ANALYTICS:

ABSTRACT:

India is generally an agricultural country. Now a days the most important emerging field in the real world is agriculture and it is the main occupation and backbone of our country. Recent developments in Information Technology for agriculture field has become an interesting research area to predict the crop yield. Crop yield prediction is the methodology to predict the yield of the crops using different parameters like rainfall, temperature, fertilizers, pesticides and other atmospheric conditions and parameters. Data techniques is very popular in the area of agriculture. Data analytics techniques are used and evaluated in agriculture for estimating the future years crop production. This paper presents a brief analysis of crop yield prediction using K-Nearest Neighbor(KNN) Algorithm for the selected region that is Mangalore, Kasargod, Hassan, Kodagu in India

SURVEY:

This paper describes and gave the details us for list of used methods, In India there are dissimilar agriculture crops production and those crops depends on the several kind of factors such as environmental science, economy and also the geographical factors covering such methodologies and methods on historic yield of dissimilar crops, it is possible to get info or data which can be supportive to farmers and government organizations for creation well decisions and for make better rules which help to increased production

This paper describes that the suggested Fresh market fruits like apples are graded into quality groups according to their size, color and shape and to the attendance of defects. This paper presents the three former points on the basis of a literature review, the research outcomes being absorbed on the last point: having extracted data from images acquired on fruits, the paper defines a classifying technique which was implemented on an existing machine and tested on Jon gold apples (bi-color fruits).

This paper discussed that the data mining extraction of unseen predictive information from huge records, is a powerful new technology with great potential to help companies focus on the most significant data in their Data warehouses. Data mining tools predict upcoming trends and performance and growth, allowing businesses to make proactive, knowledge driven decisions. This article gave us a detail project that is smearing a range of machine learning plans to problems in agriculture and horticulture.

OBJECTIVES:

The main objective is collecting agricultural dataset which can be used to analyzed for useful crop yield forecasting. To predict the crop yield with the help of data mining technique, advanced methods can be seed quality, lack of technical facilities etc the crop yield is directly influenced

Data Mining is the process of extract helpful and significant information from huge sets of data. Data Mining in agriculture field is a comparatively novel research field. Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is concerning to be expecting. In the earlier period, yield prediction was performing by considering farmer's experience on particular field and crop. In any of Data Mining actions the training data is to be collected from past data and the gathered data is used in terms of training which has to be exploited to study how to categorize future yield predictions

EXISTING SYSTEM:

Initially the raw data set was collected and it is subjected to reprocess for noise removing (replacement of missing values) and computational methods. From that dataset, it is subjected to Feature selection for make a predictive modeling. In this proposed approach it is mainly focused on Regression Techniques. Various regression analysis should be performed and it was compared and tested. Regression analysis is a form of predictive modeling technique which investigates the association between a dependent (target) and independent variable(s) (predictor). This technique is used for forecasting, time series modeling and discovers the causal effect relationship between the variables. Regression analysis indicates the significant relationships between dependent variable and independent variable and it indicates the strength of impact of multiple independent variables on a dependent variable

PROPOSED SYSTEM:

In this proposed system, the datasets are collected and refined based on the commonality. The input parameters are given. By analysing and predicting using KNN algorithm, the result are produced and some suggestions are given
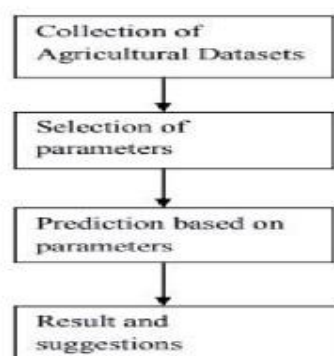

Fig. -1: Proposed Architecture

SYSTEM REQUIRMENTS:

System Requirements Specification is a document or set of documentation that describes the features and behaviour of a system or software application. It includes a variety of elements that attempts to define the intended functionality required by the users to satisfy their different users.

FUNCTIONAL REQUIRMENTS:

The Functional Requirements Definition reports and tracks the basic information expected to effectively portray business and handy necessities. The Functional Requirements Definition report is made in the midst of the Planning Phase of the endeavor. Its objective gathering is the endeavor boss, errand gathering, wander bolster, client/customer, and any accomplice whose information/respect into the necessities definitions system is require

NON-FUNCTIONAL REQUIRMENTS:

A non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. Non functional requirements specifies the quality attribute of a software system. The software system can be judged based on reliability, security, maintainability, performance, portability, scalability and flexibility

HARDWARE REQUIRMENTS:

Computer hardware specifications are technical descriptions of the computer's components and capabilities. Processor speed, model and manufacturer. Processor speed is typically indicated in gigahertz (GHz). The higher the number, the faster the computer. Random Access Memory (RAM). This is typically indicated in gigabytes (GB). The more RAM in a computer the more it can do simultaneously. Hard disk (sometimes called ROM) space. This is typically indicated in gigabytes (GB).

Processor Intel Pentium/Core – 1.7GHz and above Memory 1GB and above Storage 80GB minimum fee space Graphics 1GB and above

SOFTWARE REQUIRMENTS:

Operating System(s) Windows 7, 8, 10 Programming Language Python Framework Python

SYSYTEM IMPLEMENTATION:

UML diagram is a diagram based on the UML(Unified Modeling Language)with the purpose of visually representing a system along with its main actors ,roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system. In Fig.2 which explains the UML

diagram for our proposed system. Here crop dataset, crop recommender and users. The crop dataset will be separated into 2 datasets:Training dataset and testing dataset

Training data: A training data set is a set of data used to discover predictive relationships. Training set is used in intelligent systems, machine learning, genetic programming and statistics. Testing data: Test data is the input given to a software program. It represents data that affects or is affected by the execution of the specific module

This system works as follows: Step 1: Giving input (training data and testing data) user provide input to system as training data and testing data. Step 2: compute the prediction: Testing data and training data is given to K-NN algorithm which creates clusters from given data. K-NN algorithm gives relevant data, patterns from given datasets. Step 3: Data Acquisition: Depending on the current user location, the system mines the soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set. Step 4:Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop. The prediction is based on past production data of crops i.e.: identifying the tangible weather and soil parameters and comparing it with current conditions which will predict the crop more accurately and in a practical manner. Step 5: Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop
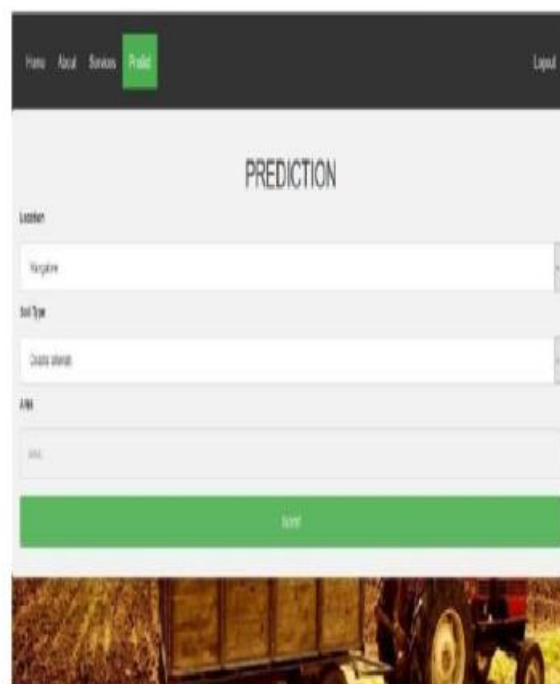
SYSTEM ARCHITECTURE:
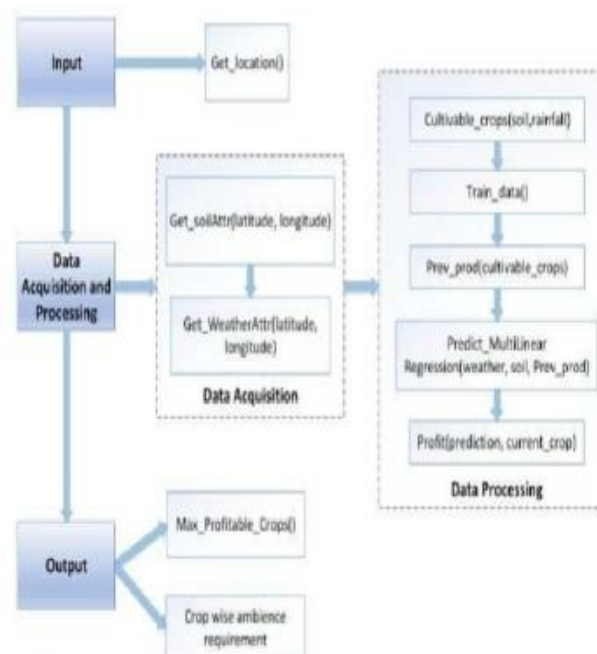
Input: The prediction of crop is dependent on numerous factors such as Soil Nutrients, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system. Data Acquisition: Depending on the current user location, the system mines the

soil properties in the respective area from the soil repository. In a similar approach, weather parameters are extracted from the weather data set. Data Processing: A crop can be cultivable only if apropos conditions are met. These include extensive parameters allied to soil and weather. These constraints are compared and the apt crops are ascertained. Multiple Linear Regression is used by the system to predict the crop Output: The most profitable crop is predicted by the system using Multiple Linear Regression algorithm and the user is provided with multiple suggestions of crop conferring to the duration of crop.

RESULT AND SCREENSHOT:

This chapter contains the snapshots of the graphical user interface(GUI)of the proposed system showing the interface, intermediate results and final result of crop prediction

CONCLUSION:

The work demonstrated the potential use of data mining techniques in predicting the crop yield based on the input parameters average rainfall and area of field. The developed webpage is user friendly and the accuracy of predictions are above 90 percent. The districts selected in the study indicating higher accuracy of prediction. The user friendly web page developed for predicting crop yield can be used by any user by providing average rainfall and area of that place. The process was adopted for all the area to improve and authenticate the validity of yield prediction which are useful for the farmers for the prediction of a specific crop.