LITERATURE SURVEY

There have been a number of research studies undertaken that focus on the importance of using data mining as a supplementary tool in transforming large volumes of agricultural data into meaningful information. Many researchers have been contributed their previous knowledge towards data mining in agriculture. There are many simulations models available for crop productivity predictions. As it depends on economic and environmental parameters so we cannot apply these existing models or methods to any other area.

D Ramesh, B Vishnu Vardhan compared statistical model Multiple Linear Regression over the Data Mining Density-based clustering technique. He concluded that comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efficiency of the [3, 7] proposed technique .

M.C.S.Geetha concluded that data mining plays a crucial role for decision making on several issues related to agriculture field. She also discusses on different data mining applications in solving the different agricultural problems .

Vrushali Bhuyar focuses on different classifier algorithms used on soil dataset to predict fertility rate. Study shows that among the classifier J48 classifier perform better to predict fertility index .

Raor ane A.A. and Kul karn i R.V., discussed few data mining techniques in their paper. They concluded that efficient technique can be developed and analyzed using the appropriate data, to solve complex agricultural problems using data mining techniques. Also recommend some of the algorithms and statistical methods that give good results in agriculture growth .

Georg Rub, Rudolf Kruse, Martin Schneider, and Peter Wagner mainly focused on Neural Network technique of data mining in their research work. They built and evaluated different networks and substantiated the assumption that the prediction accuracy of the networks rises once [10-12]more data become available at later stages into the growing season . Georg Rub also used Different regression techniques on agricultural yield data and concluded that support vector regression can serve as a better reference model for yield prediction compare to decision tree and neural network .

S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh attempted to compile the research findings of different researchers who worked on crop productivity data .

Dr. D. Ashok Kumar and N. Kannathasan focuses on different data mining techniques we can use in agriculture. Their research survey recommending that a comparison of different data mining techniques could produce an efficient algorithm for soil classification for multiple classes. The benefits of a greater understanding of soils could improve productivity in farming, maintain biodiversity, reduce reliance on fertilizers and create a better integrated soil management system for both the private and public sectors .

Anusha A. Shettar and Shanmukhappa A. Angadi used weka tool to implement data mining algorithms on dataset and concluded that: Classification algorithms like J48, LMT, LAD Tree and ID3 gives the better prediction result with respect to yield _class attribute like Poor, Good, Very_Good, and Excellent. This different accuracy result with different classification algorithms .

Yethiraj N G. concluded that there are a growing number of applications of data mining techniques in agriculture and a growing amount of data that are currently available from many resources .

Nilesh Dumbre, Omkar Chikane and Gitesh More concluded that we can say that if perfect crop recommendations are given to farmers it will definitely help to increase the crop yield and also in building the economic status of agricultural dependent countries .

Rajshekhar Borate, Rahul Ombale and his colleagues found that by using the Data Mining techniques accurate prediction of specified crop yields across different districts will help to farmers of India. They proposed a system which will recommend some inference to the farmers based on the stored data

One of the most important predictions for precision agriculture, the yield forecast is of high significance to maximise production in the mapping of returns, yield estimates, balancing crop supply with demand and crop management. Examples of ML applications include the work of an effective, inexpensive, and non-destructive meth-od that counts coffee fruits automatically on branches. The system measures the fruit of the coffee in three categories: harvestable, not harvestable and disrepair. The system also estimated the coffee fruit weight and maturity rate. This work was designed to provide coffee growers with the knowledge that would optimise their economic gains and schedule their farm work.

The developers of in which they created a computer vision method for automation during harvest, are also used for yield prediction. Even if these are not evident, the segments of the system and detect occluded cherry boughs with a complete function. The system's primary goal was to reduce manual harvesting and processing requirements.

Authors created an early yield mapping method in another study[9] to identify untimely green citrus in an outdoor citrus grove. As with all other relative research, the report aimed to provide farmers with yield-specific knowledge to help them optimise the yield and benefit growth.

In another study, the authors elaborated on ANNs and multi-temporary remote sensing data to evaluate grassland biomass (kg of dry matter /ha / day).

In another study, another study focusing on yield prediction and in particular on the prediction of wheat yield was introduced. The method developed used satellite imaging and was fused with soil data for more precise predictions.

The authors proposed a process for detecting tomatoes based on EM photographs that were taken by an aerial unmanned vehicle and remotely sensed red green blue (RGB) (UAV).

Raval et al discuss about the Knowledge Discovery Process and the basics of various Data Mining Techniques such as Association rules, Classification, Clustering, Prediction and Sequential Patterns .

Agrawal et al discuss about various Data Mining tools such as Dashboards, Text-Mining tools. They provide an overview about these tools and the various scenarios in which they can be deployed .

Grajales et al have proposed a web application that utilizes open dataset like historical production, land cover, local climate conditions and integrates them to provide easy access to the farmers. The proposed architecture mainly focuses on open source tools for the development of the application. The user can select location from map for which the details are available at one click .

Bendre et al collects data from GIS (Global Information System), GPS (Global Positioning System), VRT (Variable Rate Fertilizer) and RS (Remote sensing) and are manipulated using Map Reduce algorithm and linear regression algorithm to forecast the weather data that can be used in precision agriculture. The purpose of this study was to investigate the effective model to improve the accuracy of rainfall forecasting .

Hemageetha mainly focuses on using the soil parameters like pH, Nitrogen, moisture etc for crop yield prediction. Naive Bayes algorithm is used to classify the soil and a77% accuracy is achieved. Appriori algorithm is used to assosciate the soil with the crops that could provide maximum yield in them. A comparison of accuracy achieved during classification using Naïve Bayes, J48 and JRIP is also presented .

Rub et al presents a comparative study on the regression models that could be used for predicting yield. The algorithms discussed are Multilayer perception Model (MLP), Regtree (Regression tree), RBF (Radial Basis Function Network and SVM (Support Vector Machine). They have concluded that SVM serves as a better model as far as yield prediction is concerned .

Sujatha et al describes the purpose of various classification techniques that could be used for crop yield prediction. A few of the data mining methods, such as the Naïve Bayes, J48, random forests, support vector machines, artificial neural networks were presented. A system using climate data and crop parameters used to predict crop growth has been proposed .

Kushwaha et al predicts the suitability of a crop for a particular climatic condition and the possibilities of improving the crops quality by using weather and disease related data sets. They have proposed an analysis, classification and prediction algorithm that helps in building a decision support system for precision farming. It is based on the Hadoop file system .

Fathima et al uses data mining techniques on real time data that help in knowledge discovery. They use k means clustering algorithm to cluster the farmers based on the crop type and irrigation parameters. Appriori algorithm is used to determine which two crops are selected as a frequent item set. They generally focus on the policies that government could frame by the cropping practices of farmers .

Veenadhari et al described the purpose of data mining methods in the area of agriculture. A few of the data mining methods, such as the k-means,ID3 algorithms, the k nearest neighbor, support vector machines, artificial neural networks were presented .

Sellam et al explain the various environmental parameters like the Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI) that influences the yield of crop and the relationship among these parameters is established. Using Regression Analysis (RA), Linear Regression (LR) the various environmental factors and their infliction on crop yield is analyzed .

Raorane et al discusses the various data mining techniques for improving the crop production in agriculture. A few of Data mining methods, such as Artificial Neural Network(ANN), Decision Tree algorithm, Regression Tree, Bayesian network, Support Vector Machine(SVM), k means are used for classification .

Kaur et al analyze the different Data Mining techniques to find suitable data model that helps in achieving high accuracy for price prediction. Coimbatore market price of tomato data are collected and price is predicted using BP neural network and the result is simulated using MATLAB .

Ankalaki et al presents a comparative study on DBSCAN and AGNES algorithm for clustering. Crop yield is forecasted using MLR (Multiple Linear Regression) and a formula is derived for each crops [15]. Gayatri et al use IOT and web services to handle large amount of data. Sensors are used to collect the data and pass the data to data centre. Agriculture field images are captured and GPS is used to accurately feed the data into repositories along with their location. Far and near nodes are communicated through cloud .