International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Intelligent Computational Techniques for Crops Yield Prediction and Fertilizer Management over Big Data Environment

Sini Anna Alex, Anita Kanavalli

Abstract— Agriculture is one of the biggest fields to improve the economic rate of the country. Crop yield prediction is a new emerging idea in agriculture. There are several challenges of crops yield prediction in the field of precision agriculture are (i). Obtain minimized production due to climate change; (ii). Lead to different diseases; (iii). Availability of Water; (iv). No awareness of fertilizers and crop features; (v). Climate change; (vi). Unexpected weather events.Other loss factors in the agriculture are lowly seed quality, unplanned irrigation and exploitation of insecticides and fertilizers. The main aim of this research is to design the effective crop yield production and health risk analysis model by big data analytics model. Hence in this research our focus is on optimizing the significant parameters such as rainfall, temperature and fertilizers rate to obtain the P-values for testing the crop and also analyze the human health safety (farmers and suppliers) due to the dynamic change of environment and also soil nutrients. Big data analytics is the feasible platform to test and measure the crop grow in the particular agriculture field. It helps in climate, weather events prediction and also it is used to compute the sufficient resources for crop cultivation.

Keywords: clustering, Fertilizer planning, data analytics, prediction

I. INTRODUCTION

Agriculture is one of the biggest fields to improve the economic rate of the country. Crop yield prediction is a new emerging idea in agriculture [1], [2], [3], [4], [5]. The main aim of this research is to design the effective crop yield production and health risk analysis model by big data analytics model. prediction and also it is used to compute the sufficient resources for crop cultivation [13]. Smart farming systems are developed to address the issues of conventional farming applications. Previous applications are digital farming basis, which requires large effort to process the data from the padding field. Today smart farming system is designed by Internet of Things (IoT) and Wireless Sensor Network (WSN) [10].

Over the past few decades, machine learning algorithms have been proposed to yield crop prediction, including genetic algorithm (GA), support vector machine (SVM), linear regression, artificial neural network (ANN), Naïve Bayes (NB) and so on [14].

Revised Manuscript Received on October 30, 2019.

***** Correspondence Author

Mrs. Sini Anna Alex*, Assistant Professor in Computer Science Department of M S Ramaiah Institute of Technology.

Dr. Anita Kanavalli, Head of Computer Science Department of M S Ramaiah Institute of Technology.

Machine learning algorithms consider dynamic change of inputs. For example, dynamic changes in temperature value (if it is high), water availability, and fertilizer usage for the current time period [11]. Compared with the aforesaid machine learning algorithms, deep learning approaches have very powerful features to show the large performance in precision agriculture [9].

Deep learning algorithm is an intelligent computation approach which solves several issues, including scalability, reliability, optimization issues, and robustness.

II. LITERATURE SURVEY

Paper Title: CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture

Paper Concept: In this paper, precision agriculture is implemented for various crops yield prediction using deep learning algorithm. Classification and detection are implemented in this paper, which achieves high classification accuracy. Current deep learning algorithms have obtained 92% of classification accuracy.

Paper Title: Intelligent Agriculture and Its Key Technologies Based on Internet of Things Architecture

Paper Concept: In this study, authors have found that cluster analysis and data visualization analysis improves the agricultural product quality and improves the economic growth rate. With the functionalities such as Sensing, Identification, Transmission, Monitoring, Analysis and Feedback from the IoT sensors, processing in agriculture is effectively computed. In general, IoT sensors are used for Agriculture Induction, Identification, Monitoring for smart agriculture.

Paper Title: Application of a Genetic-Fuzzy FMEA to Rainfed Lowland Rice Production in Sarawak: Environmental, Health, and Safety Perspectives

Paper Concept: In this paper, an Improved Fuzzy Failure Mode and Effect Analysis (FMEA) with genetic algorithm are proposed for rice production. Authors of this paper have considered and resolved the health issues of farmers, environmental issues and OSHA regulations. For each farmer, the database consists of different attributes such as Gender, Age Range, and Experience in this field, Working duration per day, Health hazards by birth and working in the farming field.

Limitations

This paper considers the farmers posture (Stooping and kneeling positions), which are the root cause for health issues. In order to analyze the health effects of farmers, the usage of fertilizers and the usage in the padding field is highlight required to avoid the severity from health issues.

Time consuming task due to Fuzzy Logic and GA algorithms

Paper Title: Crop Yield Prediction Using Deep Neural Networks.

Paper Concept: This paper proposes deep neural networks (DNN) approach for crop yield production. The Syngenta crop datasets for maize yields are largely recorded i.e. 2247 locations between 2008 and 2016 and also requested recommendations to yield the performance in 2017 crop production. In order to yield the performance, DNN is used. Weather data is recorded to show the huge performance. Experiments are superior in producing the prediction accuracy, root-mean-square error (RMSE).

Limitations

There are six weather conditions are computed to yield crop production such as day length, precipitation, solar radiation, vapor pressure, maximum temperature and minimum temperature. It requires more parameters for future prediction.

Paper Title: A Smart Decision System for Digital Farming

Paper Concept: This paper shows the various potential features for smart decision making in agriculture field on the basis of real-time decisions (Irrigation variable rate, chosen parameters from the farming field and weather conditions). The parameters mined from the padding field are the index vegetation, whereas the irrigation events are flow level, pressure level, wind speed, and so on. Initially, preprocessing is initiated and processed using learning prediction rules and Drools rules. Then the smart farming system will be opened to explore the crops management and operations by the farmer information, fertilizer provider information and agricultural technicians.

Limitations

User (farmer) inserted values require dynamic update to predict the accuracy.

Paper Title: Data Mining Technology with Fuzzy Logic, Neural Networks and Machine Learning for Agriculture

Paper Concept: This paper resolves the problem arisen by farmer's decision making for agriculture applications since crop cultivation decisions will be differ according to current commodity price in the Market. Farmers are unaware of the production sustainability to improve the crop yield. Hence this paper developed the system to recommend the farmers to understand the need of current cultivation of crops in real-time basis (current requirements, dynamic environment metrics and climate change seasons and conditions) to agriculture market. For automatic identification, data mining techniques such as artificial intelligence and machine learning algorithms (fuzzy logic and neural networks) are proposed

Paper Title: CLAY-MIST: IoT-Cloud Enabled CMM index for Smart Agriculture Monitoring System

Paper Concept: In this paper, authors have presented Clay-Mist measurement (CMM) index for monitoring agriculture over cloud assisted IoT environment. There are two factors affect the crops growth, including soil moisture and temperature. Accurate levels of these factors are assessed for cropping using CMM index and then tested results are forwarded to farmers. Finally farmer takes the remedy for health growth of crops.

Limitations

Computational expensive is high due to CMM index measurement

Paper Title: Modeling a Predictive Analytics Methodology for Forecasting Rice Variety and Quality on Yield On Farm and Farming Attributes using Big Data

Paper Concept: Authors of this paper have investigated the diverse farm attributes with their comfort assess levels and also cropping methods to maximize the growth rate on large volume of dataset. For experimentation, rice crop is tested in Kanchipuram, India. Big data clustering approach is proposed in this paper to predictive analytics.

Paper Title: Soil based Fertilizer Recommendation System using Internet of Things

Paper Concept: In this paper soil based fertilizers recommendation application is implemented in IoT. Soils are differing based on its colour, water irrigation, and cultivation. Hence this paper suggests fertilizers to crop growth based on the soil type. Naïve classifier is used in this paper to test the soil characteristics, which is a simple probabilistic model for soils testing, but it shows the strong dependency between the feature vectors.

Limitations

In naïve bayes algorithm, there is an assumption (Class-Conditional Independence), which reduces the prediction accuracy and it does not work well when the features are highly correlated with each other.

III. RELATED WORK

Analysis of agriculture data using data mining techniques: Application of Big Data

Paper Concept: In this paper, authors have proposed data mining (clustering) techniquessuch as PAM, DBSCAN, Multiple Linear Regression and CLARA. The input data is collected for 6 years with the following entities such as State-Karnataka (28 Districts), Crops (Cotton, Groundnut, Rice, Wheat, and Jowar), Seasons (Rabi, Summer, and Kharif), Area (in Hectares), Protuction (in ton), Average Temperature (℃), Average Rainfall (mm), Soil PH value, Soil Type, Major Fertilizers, Nitrogen (kg/Ha), Phosphorous (Kg/Ha), minimum rainfall and temperature required, and potassium (Kg/Ha). For clustering similar group of data (rain fall, temperature, and soil type), modified DBSCAN method is used. Finally, authors have proved that the DBSCAN was superior than other clustering approaches.

Problems Defined

- DBSCAN requires more time when form clusters for large volume of data an also KNN plot is used to find the optimal Eps value, which is a time consuming task
- Multiple linear regression find the optimal parameters (optimal temperature, worst temperature and rainfall) to yield crop production, which does not sufficient, and hence proper soil nutrients, irrigation of water, fertilizers type and usage are required

- Data preprocessing steps are required, which improves the crop yield production accuracy

Proposed Solutions

- We proposed Nearest Neighbor-based MapReduce algorithm, which is improves cluster formation speed and it supported large volume of data.
- We propose deep learning approach known as Convolutional Neural Network to find the optimal parameters and determine the current stage of crop, which maximize the crop yield production at farmer size.
- Before the classification process, we present the data enrichment process where we enrich the inbound data that found in the farm field.

A Granular GA-SVM for Big Data in Agricultural Cyber-Physical Systems

Paper Concept: In this paper authors have proposed Genetic Algorithm (GA) and Support Vector Machine (SVM). Three methods are proposed to minimize the large granule into small i.e. Min-Median-Max,Quartile-Median granulation, and Fuzzy granulation. In integrated GA-SVM, GA is proposed to find the optimal parameters of SVM such as Penalty and Kernel. This model is applied for big data analytics and management and it is superior in terms of prediction accuracy and computation time. There are 6 parameters are considered for agricultural applications such as air temperature, air humidity, soil temperature, soil humidity, wind speed and wind direction. In this paper, precision agriculture is presented for Apple.

Problems Defined

- Granulation methods does not supported for complex data. In addition, data canonicalization process is required to improve the crop yield production.
- For large complex data, the integrated GA and SVM is not sufficient

Proposed Solutions

- For classification, we propose CNN with Jaccard Similarity function, which finds the clusters purity and then the classification is done.
- Data canonicalization process is proposed which implemented using Symmetry Matching Canonicalizerfor representing the records in the collected set. This matching of canonicalizer maximizes the searching speed and minimizes the search space.

Precision Agriculture for Banana using Wireless Sensor Network.

Paper Concept: This paper presents the study of banana precision agriculture in wireless sensor network. The optimal parameters are computed according to the irrigation water, geographical region and soil type. This paper handles the unexpected weather events change issues, and provide the step to improve the economic growth of Indian farmers for precision agriculture. This paper alerts the farmers to avoid Sigatoka disease.

Problems Defined

- Authors of this paper have not proposed the monitoring method for precision agriculture, which is required to update the dynamic parameters change.
- Direct use of sensed data from farming field leads to misclassification issues, which requires some processing to analyze the data.

Proposed Solutions

- We propose the CNN with Jaccard similarity architecture and it is used for accurate prediction
- We considered Data enrichment strategy to rectify the errors on inbound data.

Prediction of Major Agricultural Fruits Production in Pakistan by using an Econometric Analysis and Machine Learning Technique

Paper Concept: In this paper authors have proposed the econometric analysis for primary fruits prediction using machine learning approach. Various fruits data is collected from 1980 to 2015, which gathers on the basis of time-series sequences. The machine learning algorithm is used to predict the crop yields for future prediction on Pakistan. Ordinary least square method and augmented dickey fuller are used to analyze and test the results. Linear regression model is used to identify the future by using the set of fruits data.

Problems Defined

- Future prediction using linear regression does not supported for non-linear data. However it considers several assumptions at the initialization stage. If the assumption is wrong for non-linear data, it tends to misclassification. Linear regression is a simple model that does not suited for large volume of complex data.

Proposed Solutions

- We propose deep learning approach known as Convolutional Neural Network to find the optimal parameters and determine the current stage of crop, which maximize the crop yield production at farmer size.

Artificial Neural Network Modelling in the Prediction of Banana's Harvest

Paper Concept: This paper proposes artificial neural network for crop (banana) yield prediction. Data were collected from 2011-2017 i.e. six years are collected for experiment banana crop. These experiments considered fertilization, soil preparation and correction for analysis of banana crop. In addition, some meteorological data according to Temperature, Humidity and Precipitation are acquired from INMET (National Institute of Meteorology).

Problems Defined

- Lack of consideration for fertilization, and banana harvesting prediction using ANN may leads to uncertainty issues.

Proposed Solutions

- We considered fertilizer data for increasing crop yields to current years

IV. PROPOSED WORK

Our designed problem statement is to build big data analytics model for precision agriculture and fertilizer management to improve the agriculture field, which improve the economic growth of the country. Our overall problem statement is visualized in Figure 4.1 as pictorial representation.

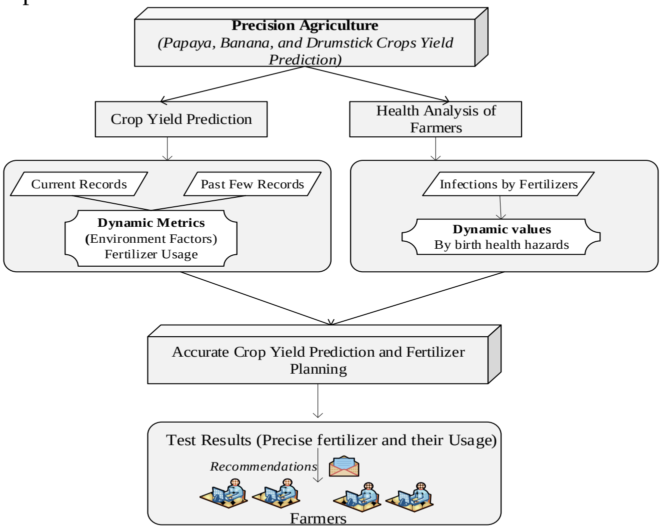


Figure 4.1 Overall design of the precision agriculture big data analysis model

OBJECTIVES

- Predict the crop (Papaya, Drumstick, and Banana) yield and effective fertilizer usage by adding soy milk for the next 5 years (2018-2023)based on the average rainfall, soil moisture, soil nutrients etc.
- Health analysis for the farmers who use fertilizers and the suppliers who supply fertilizers.Predict the health hazards that may happen with respect to the usage of excess fertilizers.

In our proposed work, we address the problems stated in the problem definition section. We propose the intelligent computational approaches for crop yield prediction and fertilizer management over big data environment. We collect the data, including Nitrogen (N), Phosphorous (P), Calcium (C), Magnesium (M), and Sulphur (S). Our aim and objectives are achieved by the following phases:

(i). Phase 1: Data Enrichments

(ii). Phase 2: Data Clustering

(iii). Phase 3: Data Classification

(iv). Phase 4: Recommendations

Each phase is detailed in the following

(i). Phase 1: Data Enrichments

In this we pre-process the data, which is collected from farming field. In general, inbounded data does not loaded into the system for maximizing prediction accuracy and minimizes the system overhead to predict the crops. There are four steps are primarily implemented to enrich the data quality, including Data Deduplication, Data Denoising, Data Canonicalization, and Component Breakdown. In first step, repeated values are removed by estimating the similar values in the stored data. In second step, noise values are removed by using Locally Linear Embedding technique (LLE), which reduces the data dimensionality. In third step, data canonicalization is implemented by Symmetry Matching Canonicalizer algorithm. A smart farming system creates the same entity for different type of texts. For e.gAbc,' 'ABC+,', 'ABC, 10' represents the same pesticide. All the entries are enrolled in a single form. Finally in the fourth step, component breakdown is implemented to understand the fertilizers nutrients by the name. For e.g. "10-10-25-45" can be expanded into N (Nitrogen) [10%], P2O5 (Phosphorus pentoxide) [0%], K2O (Potassium oxide) [26%], and SO3 (Sulphur trioxide) [6%].

(ii). Phase 2: Data Clustering

This step helps to analyze the data easier and faster. In data clustering, we group the similar data points into one cluster, which is performed by the Nearest Neighbor-based MapReduce clustering algorithm. In this step, we process the following:

- Split the input data into multiple partitions
- Distribute them to available nodes in the HDFS (Hadoop Distributed File System)
- Mapper function is executed first for concurrently cluster each data partitions, and then find the same boundary data points assigned to various partitions
- Apply Reducer function to get the clustering IDs to form the clusters

Then we compute the cluster purity using Jaccard Similarity Function as shown in Figure 5.2.

(iii). Phase 3: Data Classification

We perform data classification based on clustered data using Convolutional Neural Networks (CNN), which consists of several layers including Convolutional Layers, Max_pooling Layers, Fully Connected Layers, and Softmax Layer as shown in Figure 3.2. In softmax layer, we compute the current crop condition and then classified into three classes: (i). Healthy growing, (ii). Alert forwarder require, and (3). Alert for emergency. This step is accomplished on the basis of optimal parameters. For input parameters, we considered the following parameters in CNN.

(1). Temperature

(2). Rainfall

(3). PH

(4). Water level

(5). N-nitrogen

(6). P-phosphorous

(7). K-potassium

(8). C-calcium

(9). M-magnesium

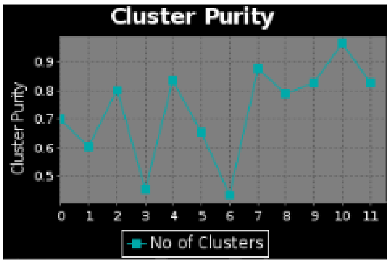
(10). S-sulphur

(iv). Phase 4: Recommendations

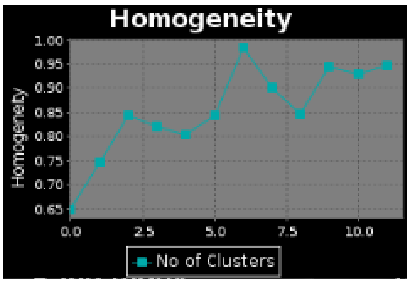
If the crop is class 1 or class 2, the tested result is sent to the farmer via Mail. And also this test result is forwarded to Agronomist, who find the P-value by optimum parameters which are selected. Agronomist is finding the optimal conditions using Deep Reinforcement Learning algorithm (State, Action and Rewards). Based on the current conditions, dynamic parameters of environment are updated concurrently. Then this result is forwarded to the Farmer for yield crop performance, and avoids crops failure. Agronomist suggests the optimal conditions, fertilizer type and the usage, based on the metadata values as shown in Table 3.1 and stored in the MongoDB.

The experiments are conducted and the performance is compared with the Modified DBSCAN (Reference 1) for the following metrics.

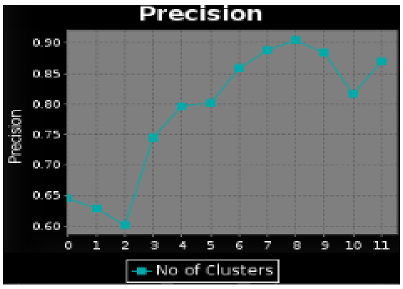
- Cluster Purity (Range 0-1)



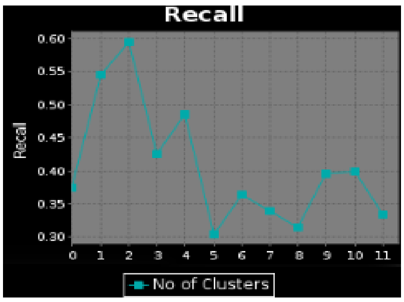
- Homogeneity (Range 0-1)



- Completeness (Range 0-1)
- Precision (Range 0-1)



- Recall (Range 0-1)



- Prediction Accuracy (Range 0-1)
- F-measure (Range 0-1)

- Rand Index (Range 0-1)
- Error Rate (Range 0-1)
- Execution Time (S)

System Architecture

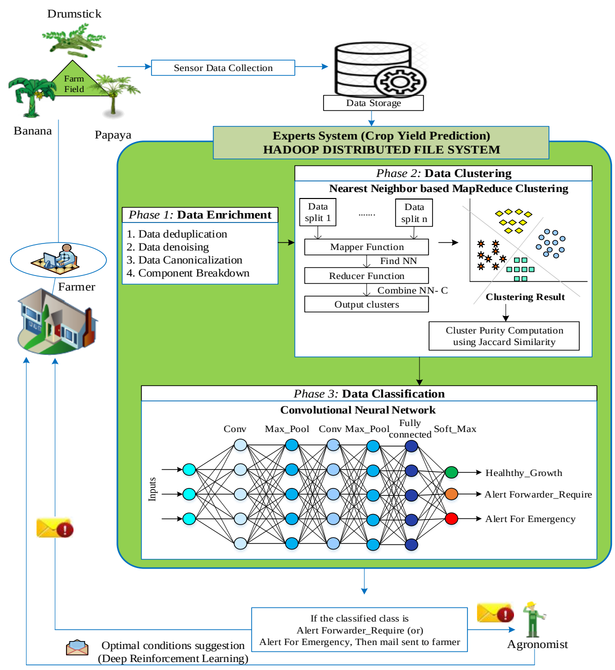


Figure 4.2 The proposed system architecture

V. CONCLUSION AND FUTURE SCOPE

The data can be trained and analyzed using R Tool effectively. Predict future outcomes and forecast as per market situations using Predictive Analytics for the proposed model can be done. In this paper we have designed the model based on Hadoop Distributed File system.

Using Convolutional neural network the data classification is done ,theclassified class can share and alert farmers through any web interface.

ACKNOWLEDGEMENTS

We acknowledge the management and staff of M S Ramaiah Institute of Technology for their support and encouragement to do this research.

REFERENCES

1. Majumdar, J., Naraseeyappa, S., &Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data, 4(1).
2. Ruan, J., Jiang, H.Q., Li, X., Shi, Y., Chan, F.T., &Rao, W. (2019). A Granular GA-SVM Predictor for Big Data in Agricultural Cyber-Physical Systems.
3. Pujari, S., &Bogiri, N. (2017). Precision Agriculture for Banana Using Wireless Sensor Network. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).
4. Rehman, A., Deyuan, Z., Hussain, I., Iqbal, M. S., Yang, Y., &Jingdong, L. (2018). Prediction of Major Agricultural Fruits Production in Pakistan by Using an Econometric Analysis and Machine Learning Technique. International Journal of Fruit Science, 1–17.
5. De Souza, A. V., BoniniNeto, A., Cabrera Piazentin, J., Dainese Junior, B. J., Perin Gomes, E., dos Santos Batista Bonini, C., & Ferrari Putti, F. (2019). Artificial neural network modelling in the prediction of bananas' harvest. ScientiaHorticulturae, 257, 108724.
6. Zheng, Y., Kong, J., Jin, X., Wang, X., Su, T., &Zuo, M. (2019). CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. Sensors.
7. Chen, J., & Yang, A. (2019). Intelligent Agriculture and Its Key Technologies Based on Internet of Things Architecture. IEEE Access, 7, 77134–77141
8. Sang, A. J., Tay, K. M., Lim, C. P., &Nahavandi, S. (2018). Application of a Genetic-Fuzzy FMEA to Rainfed Lowland Rice Production in Sarawak: Environmental, Health, and Safety Perspectives. IEEE Access, 6, 74628–74647.
9. Khaki, Saeed& Wang, Lizhi. (2019). Crop Yield Prediction Using Deep Neural Networks. Frontiers in Plant Science. 10.
10. Cambra, Carlos &Sendra, Sandra &Lloret, Jaime &Tomás, Jesús. (2019). A Smart Decision System for Digital Farming. Agronomy. 9. 216.
11. Kale, S. S., &Patil, P. S. (2018). Data Mining Technology with Fuzzy Logic, Neural Networks and Machine Learning for Agriculture. Advances in Intelligent Systems and Computing, 79–87
12. Mekala, M. S., &Viswanathan, P. (2018). CLAY-MIST: IoT-Cloud Enabled CMM index for Smart Agriculture Monitoring System. Measurement, 134, 236-244
13. SenthilVadivu, Prasanna Devi, VinuKiran, Manivanna, "Modeling a Predictive Analytics Methodology For Forecasting Rice Variety and Quality on Yield on Farm and Farming Attributes using Big data", International Journal of Pure and Applied Mathematics, 116(5), 61-65, 2017
14. KomalBodake, RutujaGhate, HimanshiDoshi, PriyankaJadhav, and BalasahebTarle, "Soil based Fertilizer Recommendation System using Internet of Things", MVP Journal of Engineering Sciences, 1(1), 1-7, 2018

AUTHORS PROFILE

Mrs. Sini Anna Alex is working as an Assistant Professor in Computer Science Department of M S Ramaiah Institute of Technology. Her areas of interest include, ubiquitous computing, compiler design, mobile and wireless sensor networks, distributed, database systems and algorithm analysis. She published her papers in Scopus conferences and journals. She received Venus International Women Award for Young Woman in Engineering - 2018. She received RULA award for the best paper on pervasive computing systems.

Dr. Anita Kanavalli is working as the Head of Computer Science Department of M S Ramaiah Institute of Technology. Her areas of interest include Adhoc networks, high performance computing and microprocessors. She has membership in ISTE and IEEE. Her some of the research papers are published in Scopus Indexed Journals and Conferences. She completed certification course in CCNA and subject expert for EDUSAT.