

# Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning

<sup>1</sup>Girish L, <sup>2</sup>Gangadhar S, <sup>3</sup>Bharath T R, <sup>4</sup>Balaji K S, <sup>5</sup>Abhishek K T

<sup>1</sup> Assistant Professor, <sup>2,3,4,5</sup> UG students, Department of CSE, CIT Gubbi Tumkur, Visvesvaraya Technological University, Belagavi, Karnataka, India

<sup>2</sup> gangadhar.smg24@gmail.com

**Abstract**— Smart Agriculture is a development that emphasizes the use of information technology in the farming. Most of the population in India depending on agriculture. This situation is one of the reason, that hindering the development of country. Nowadays, even though farmers get more yield for their crop but the market price for that crop will be less, in that case farmers get loss for their product and vice versa. Particularly, when growing new crops, farmers face the risks of both market price and production problems. To overcome these problems, a machine learning technology is used. Predictive analysis is a branch of data mining which predicts the future probabilities and trends. The prediction will help the farmers to choose whether the particular crop is suitable for specific rainfall and crop price values. This approach is to increase the net yield rate of the crop, based on rainfall. Prediction can be carried out by using various machine learning algorithms like linear regression, SVM, K NN method and decision tree algorithm out of which SVM is giving the highest efficiency. The predictive analysis technique can be implemented in several government sectors like APMC, kissan call center etc., by which the government and farmers can get the information of the future rainfall, crop yield and the market price.

**Keywords**—Machine learning; Predictive analysis; Support Vector Machine (SVM);

## I. INTRODUCTION

Agriculture gave birth to civilization in India. India is an agrarian country and its economy largely based upon crop productivity. Thus agriculture is the backbone of all business in India. Our aim is to increase the net yield rate of the crop, based on rainfall. This helps use to predict the crop yield and suggest the best crop thereby improving the quality and profitability of the agriculture sector by processing the datasets. This prediction will help the farmers to choose whether the particular crop is suitable for specific rainfall and crop price values. This prediction can be carried out by using various machine learning algorithms like linear regression, support vector machine, K-NN method and decision tree where high accuracy and speed can be achieved [4].

Nowadays if the farmers get more yield for their crop but the market price for that crop will be less, in that case farmers get loss for their product and in vice versa if market price is more but crop yield will be less and poor quality, for that situation also they get less profit. Particularly when growing new crops, farmers face the risks of both market price and production problems. Day by day technologies are improving even though there are no useful technologies to help the farmers about market price, crop production and

rain fall prediction information thus they fail in their production and also in marketing.

The main objective of this paper is to develop an application that suggest the crops for farmers in Tumakuru district based on predictive analysis. This paper is aiming to assist the farmer in selecting the crops based on predicted rainfall values [6].

## II. RELATED WORK

### A. Machine Learning Technique

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Under Artificial Intelligence machine learning is the main area where a machine works like an intelligent human brain. Machine has to initially think and then learn like a brilliant man. As like a person in society learns from experiences and former data that it is exhibit to and according to that the machine takes decisions in upcoming events [2].

There are some methods of algorithms and learning in machine learning:

1. **Supervised Learning:** In this method of machine learning few training dataset are given to the particular

algorithm on the support of these datasets machine can analyze the input values and their output values.

2. **Unsupervised Learning:** In this method no output values are given to the algorithm, this method of learning is considered as unsupervised learning. We do not have any outcome value to estimate or predict.

3. **Reinforcement Learning:** To make specific decisions, the machine is trained using this algorithm. In this method, the machine is trained continuously using trial and error methods by exposing it to the past dataset environment.

**B. Literature work**

Author Yun HwanKim et al [4], developed a Crop Pest Prediction Method using Regression and Machine Learning Technology. This technique can help farmers to reduce damages of crop and increase their income.

Author Chandrashekhar Azad et al [5], proposed Design and Analysis of Data mining based Prediction model for Parkinson’s disease. In this paper prediction of speech articulation symptom is done easily by allowing the disease affected people for formulation and analysis.

Author Perpetua Noronhal et al [6], states about Comparative Study of Data Mining Techniques in Crop Yield Prediction. In this, Bi-clustering Technique has major Scope to analyze but it has been rarely used in this field. Bi-clustering technique can use wisely whenever necessary.

Author Sanjay D. Sawaitul et al [7], briefed Classification and Prediction of Future Weather by Using Back Propagation Algorithm. The forecasting system can be extended to note the reading from cyclone data and also to increase the range of wireless devices or sensors used for transmitting the data to the pc. The system can be extended for prediction of large areas at a time.

**III. PROPOSED SYSTEM**

There are multiple ways to increase and improve the crop yield and quality of the crops. This paper proposes an idea to identify the suitability of crops for a particular soil which is based on the basic assets. We use various predictive algorithms where accuracy can also be measured.

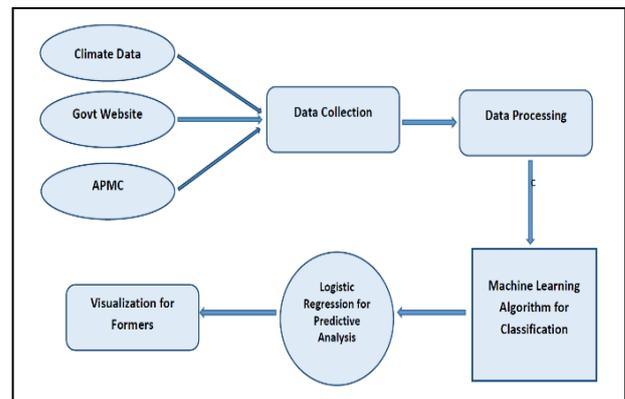
For analyzing and predicting the suitable crop for a particular season.

- The proposed system will gives the best analysis using prediction system and this can be used for crop disease prediction and crop yield production to farmers in their agriculture.
- In real time it can be used in APMC for real efficient use. In future planned to develop mobile application this can be accessed in remote areas by farmers.
- Selection of the appropriate crop plays a major role in maximizing the crop yield. For crop selection and proper yield prediction machine learning gives effective algorithm. Many techniques like K-nearest

neighbors and Support Vector Method (SVM), linear regression, decision tree are used for the prediction.

The below Fig. 1. shows that the architecture of system.

Fig. 1. System Architecture



**Data collection:** It is the well-ordered method for collecting and measuring data from different fields to get approximate and full image of required area.

**Data pre-processing:** It is a data mining process that involves cleaning the needed data, integration the data, data transformation, reduction of selected data and discretization of dataset as depicted in the Fig. 2.

**Classification:** In machine learning, classification is a method in which prediction of the training dataset group is done and are used to recognize the class names on generated dataset.

**Logistic regression:** Regression is a process of learning a function that categorize a dataset to a predicted values.

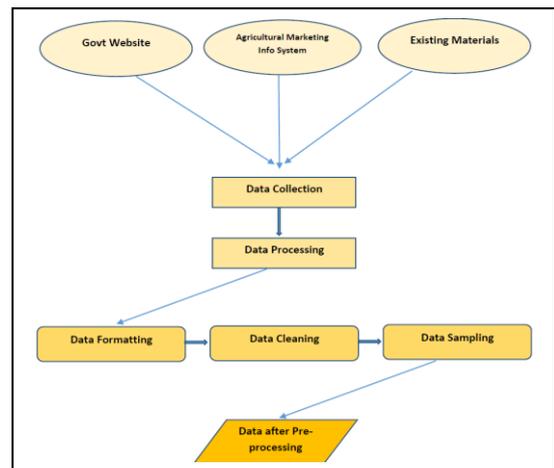


Fig. 2. Flowchart for data collection and pre-processing

**Predictive analysis:** Predictive analysis is a branch of data mining which predicts the future probabilities and trends. Plans and strategies are carried out that improve outcomes. Predictor is central element in predictive analysis, a mutable which can be measured to a distinct or additional entity to forecast future behaviour [3].

Soil characteristics can be studied using several data mining techniques. For instance, the k-means algorithm is used for soil classification in the mixture of GPS-based method and

plants same as support vector machines (SVM) to classify crops. ID3 algorithm is used to generate the decision tree from a datasets. The k-nearest neighbour (KNN) reserves all available cases and categorise new cases based on a parallel measure and also used in statistical prediction and pattern recognition. The multidisciplinary method of integrating computer science with Indian agriculture is helpful in estimating agriculture crops successfully[1].

#### IV. PREDICTION ALGORITHM

There are various methods for prediction, which includes:

- Linear Regression
- SVM
- KNN
- Decision Tree

**Linear Regression:** Linear regression was the first method in regression analysis which can be studied rigorously, and to be used widely in practical application. This is because models which are non-linearly related to their parameters are difficult to fit than models which depend linearly on their unknown parameters.

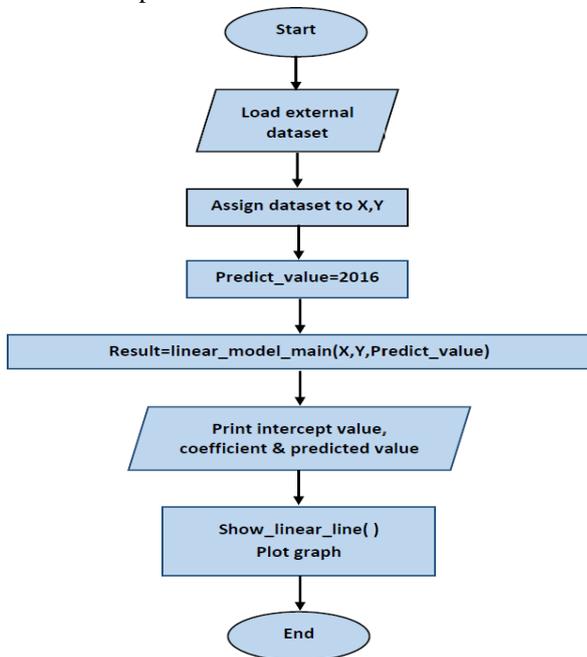


Fig. 3. Flowchart for linear regression model

TABLE I. TEST CASES FOR LINEAR REGRESSION METHOD

| SI No | Input |               | Expected Output Range | Predicted Output | Comment      |
|-------|-------|---------------|-----------------------|------------------|--------------|
|       | Year  | Rainfall (mm) |                       |                  |              |
| 1     | 2012  | 473           | 450-650               | 625              | 75%efficient |
| 2     | 2013  | 584           | 450-650               | 629              | 93%efficient |
| 3     | 2014  | 690           | 450-650               | 633              | 92%efficient |
| 4     | 2015  | 511           | 450-650               | 637              | 80%efficient |
| 5     | 2016  | 515           | 450-650               | 641              | 80%efficient |

**Support Vector Machine:** SVM can prioritize the correctness of classifying while maximizing the margin, but also perform as other learning algorithm. If there is a point that can't be classified correctly, in that case retain its largest margin, SVM will treat it as outliers and can securely ignore the points [7].

TABLE II. TEST CASES FOR SVM METHOD

| SI No | Input |               | Expected Output Range | Predicted Output | Comment      |
|-------|-------|---------------|-----------------------|------------------|--------------|
|       | Year  | Rainfall (mm) |                       |                  |              |
| 1     | 2012  | 473           | 450-650               | 544              | 87%efficient |
| 2     | 2013  | 584           | 450-650               | 583              | 99%efficient |
| 3     | 2014  | 690           | 450-650               | 585              | 84%efficient |
| 4     | 2015  | 511           | 450-650               | 512              | 99%efficient |
| 5     | 2016  | 515           | 450-650               | 417              | 81%efficient |

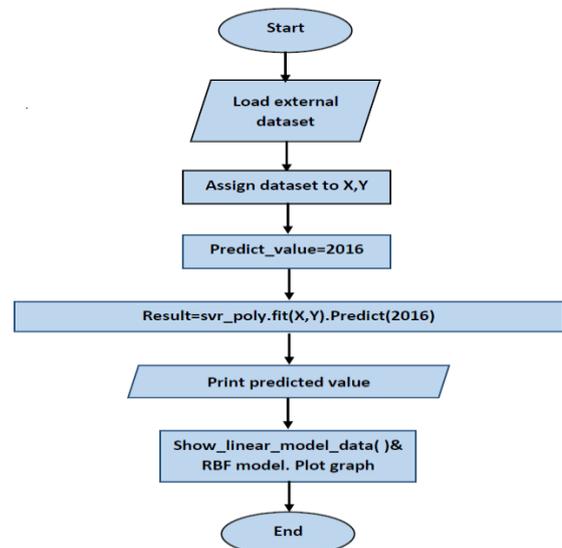


Fig. 4. Flowchart for SVM model

**K-Nearest Neighbors:** K-nearest neighbors is a simple algorithm which can be used for classification and regression. In statistical estimation and pattern recognition a KNN has been used as a nonparametric technique. In both classification and regression cases, the input consist of k closest training examples in the feature space and output depends on whether k-NN is used for classification or regression.

TABLE III. TEST CASES FOR K-NN METHOD

| SI No | Input |               | Expected Output Range | Predicted Output | Comment      |
|-------|-------|---------------|-----------------------|------------------|--------------|
|       | Year  | Rainfall (mm) |                       |                  |              |
| 1     | 2012  | 473           | 450-650               | 627              | 75%efficient |
| 2     | 2013  | 584           | 450-650               | 561              | 96%efficient |
| 3     | 2014  | 690           | 450-650               | 595              | 86%efficient |
| 4     | 2015  | 511           | 450-650               | 564              | 91%efficient |
| 5     | 2016  | 515           | 450-650               | 600              | 86%efficient |

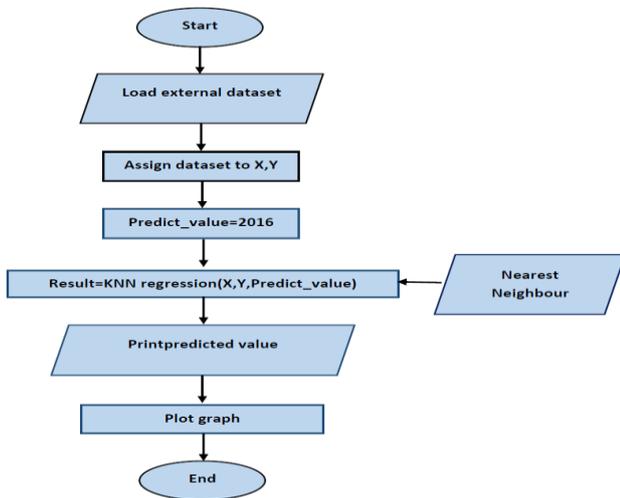


Fig. 5. Flowchart for K-NN model

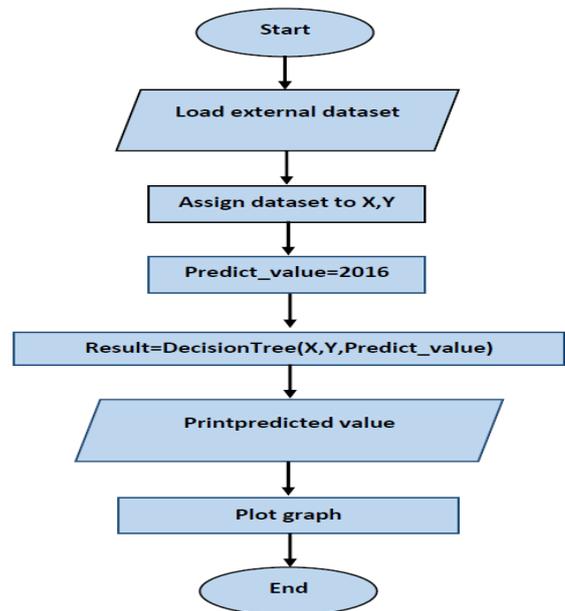


Fig. 5. Flowchart for decision tree model

**Decision Tree:** Decision tree is a supervised learning method used for classification and regression.

The main aim is to create a model that predicts the value of a targeted variable by learning simple decision rules indirect from the data features [8].

Various Decision tree are algorithms are ID3, C4.5, C5.0 and CART

- ID3- It is a multiway tree, finding for each node of the largest information for categorical targets gained by yielding the categorical features.
- C4.5- It is successor to ID3. C4.5 converts trained trees into set of if-then rules and each rule accuracy is then evaluated to verify the order in which they should be applied.
- C5.0- This algorithm uses less memory and make rule sets which is smaller than C4.5 while being more precise.
- CART- It exactly look like C4.5 but differ in numerical target variable supporting and compute rule set is not supported.

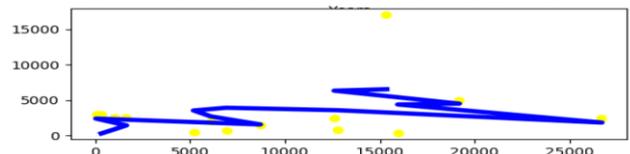
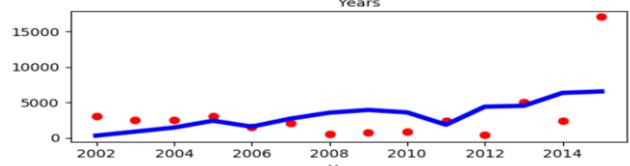
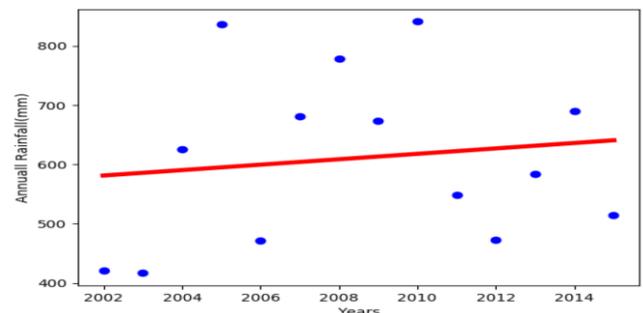
TABLE IV. TEST CASES FOR DECISION TREE METHOD

| Sl No | Input |               | Expected Output Range | Predicted Output | Comment       |
|-------|-------|---------------|-----------------------|------------------|---------------|
|       | Year  | Rainfall (mm) |                       |                  |               |
| 1     | 2012  | 473           | 450-650               | 642              | 73% efficient |
| 2     | 2013  | 584           | 450-650               | 690              | 84% efficient |
| 3     | 2014  | 690           | 450-650               | 637              | 92% efficient |
| 4     | 2015  | 511           | 450-650               | 561              | 91% efficient |
| 5     | 2016  | 515           | 450-650               | 511              | 99% efficient |

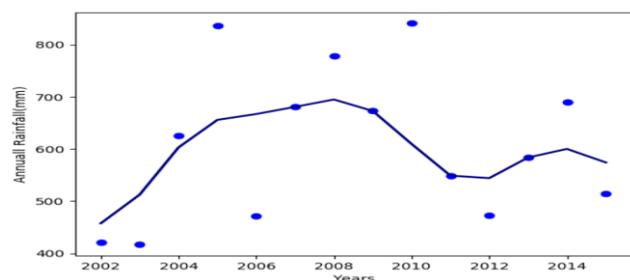
### V. EXPERIMENTAL RESULTS

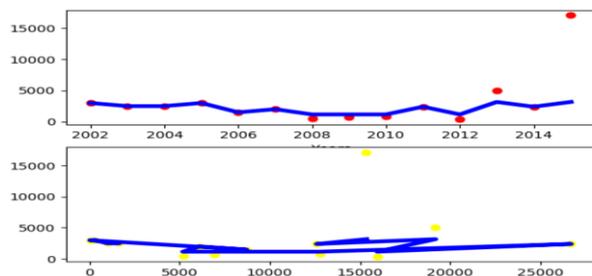
This section contains the modules linear regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree results in detail.

#### Linear Regression:

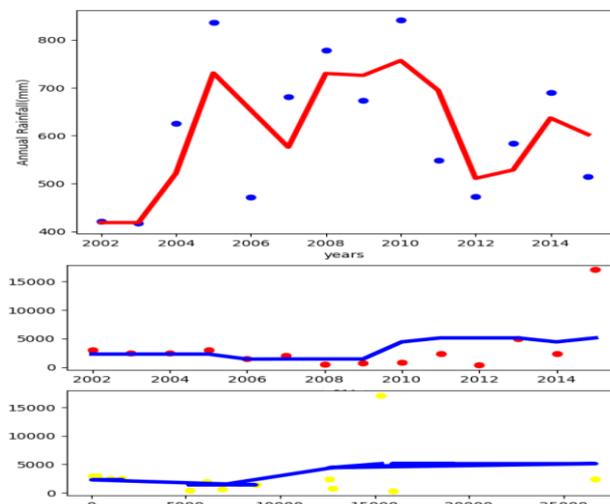


#### Support Vector Machine:

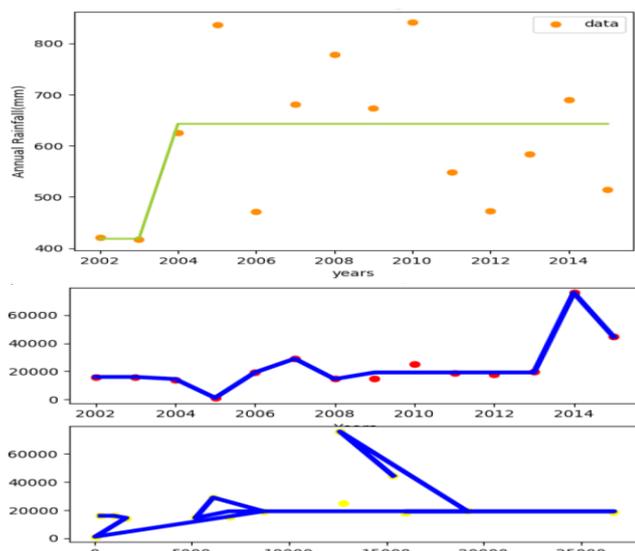




**K-Nearest Neighbor:**



**Decision Tree:**



From the above experimental outputs, it is observed that Support Vector Machine algorithm is giving the highest efficiency compared to all other algorithms.

TABLE V. ALGORITHM EFFICIENCY TABLE FOR RAINFALL DATA

| Algorithm         | Input |               | Expected Output Range | Predicted Output | Efficiency    |
|-------------------|-------|---------------|-----------------------|------------------|---------------|
|                   | Year  | Rainfall (mm) |                       |                  |               |
| SVM               | 2016  | 515           | 450-650               | 551              | 93% efficient |
| K-NN              | 2016  | 515           | 450-650               | 602              | 85% efficient |
| Linear regression | 2016  | 515           | 450-650               | 645              | 79% efficient |
| Decision tree     | 2016  | 515           | 450-650               | 637              | 80% efficient |

TABLE VI. ALGORITHM EFFICIENCY TABLE FOR MARKET DATA

| Algorithm         | Input |                   | Expected Output Range |             | Predicted Output |       | Efficiency    |
|-------------------|-------|-------------------|-----------------------|-------------|------------------|-------|---------------|
|                   | Year  | Arrival (quintal) | Min                   | Max         | Min              | Max   |               |
| SVM               | 2016  | 21948             | 500-17500             | 11000-50000 | 2162             | 19563 | 89% efficient |
| K-NN              | 2016  | 21948             | 500-17500             | 11000-50000 | 6067             | 41848 | 52% efficient |
| Linear regression | 2016  | 21948             | 500-17500             | 11000-50000 | 3700             | 19000 | 86% efficient |
| Decision tree     | 2016  | 21948             | 500-17500             | 11000-50000 | 17100            | 44500 | 49% efficient |

**VI. CONCLUSION AND FUTURE ENHANCEMENT**

In this paper, we have performed the predictive analysis using real time data collected from agricultural department and government website. By applying predictive analysis on the collected data, we can help the farmers to choose whether the particular crop is suitable for specific rainfall and crop price values.

In future enhancement, using deep learning technology and image processing technique we will implement for crop disease prediction [5] by collecting images of leaves, crops etc., which still improves the net yield of crop production.

**ACKNOWLEDGMENT**

We acknowledge and express our sincere thanks to Dr. Shantala C.P, Vice Principal and HOD, Department of CSE, CIT Gubbi, Tumkur for giving us an opportunity to expose ourselves into such research and for her many valuable suggestions and continued encouragement and support in this paper endeavors.

**REFERENCES**

- [1] Ramya, M., Chetan Balaji, and L. Girish. "Environment Change Prediction to Adapt Climate-Smart Agriculture Using Big Data Analytics." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 (2015).
- [2] Girish, L., and Sridhar KN Rao. "Mathematical tools and methods for analysis of SDN: A comprehensive survey." Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on. IEEE, 2016.
- [3] Sahana, D. S., and L. Girish. "Automatic Drug Reaction Detection Using Sentimental Analysis." , International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015
- [4] Yun HwanKim, Seong JoonYoo, Yeong HyeonGu, Jin HeeLim, DongilHan and Sung WookBaik. "Crop Pest Prediction Method using Regression and Machine Learning Technology".
- [5] Chandrashekhar Azad, Sanjay Jain, Vijay Kumar Jha. "Design and Analysis of Data mining based Prediction model for Parkinson's Disease". International Journal of Computer Science Engineering (IJCSE) Vol. 3 No.03 May 2014.
- [6] Perpetua Noronha1, Divya J, Shruthi B.S. "Comparative Study of Data Mining Techniques in Crop Yield Prediction". International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) Vol. 5, Special Issue 2, October 2016.
- [7] Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr. P. N. Chatur. "Classification and Prediction of Future Weather By Using Back Propagation Algorithm". International Journal of Emerging Technology and Advanced Engineering Volume 2, Issue 1, January 2012.
- [8] Rohit Punnoose, Pankaj Ajit "Prediction of Employee Turnover in Organization using Machine Learning Algorithms". International Journal of Advanced Research in Artificial Intelligence (IJARAI), Vol. 5, No. 9, 2016.