



# Two Stage Classification Model for Crop Disease Prediction

<sup>1</sup>Anuradha, <sup>2</sup>Kuldeep Kaswan, <sup>3</sup>Sugandha Singh

<sup>1</sup>M.Tech Scholar, CE, PDMCE, Bahadurgarh

<sup>2</sup>Assistant Professor, CSE, PDMCE, Bahadurgarh

<sup>3</sup>Associate Professor, PDMCE, Bahadurgarh

*Abstract- To acquire the knowledge or the extracted information there are number of classification models. In this paper, an improved hybrid model is presented. The work is here defined in the integration of HMM and neural network models. The HMM is here defined as the feature extraction approach using which the weights are assigned to the information set. In second stage, fuzzy adaptive neural is applied to perform the disease prediction. The obtained shows that the work has provided the high recognition rate.*

*Keywords: Data, DM techniques, WSN, Machine Learning, DM parameters.*

## 1. Introduction

### Data

Data mining is all about data. Data may be in the form of text, image, audio, facts, numbers etc. Today data are in many forms and in different databases like operational, nonoperational, Meta data etc.

### Information

There are many data which provide information like patterns, association etc. Mined information represents the semantic model of datasets. This model is used for future data and also helps in future prediction and classification of new data Data mining is a science.

## Knowledge

Knowledge is summary information about the data. Knowledge can be in form of patterns which are used for future directions.

## Mining data and not mining data

When u put a query on web search engine e.g.-education then web search engine returned to edu.com, edu.study etc. this is done according to the context of query. This data are mining data and not mining data are like put a query on www for information about education.

## Data mining

Data mining is a computation AI process which involves methods like database and data warehouse, neural network, artificial intelligence and machine learning[12]. Data mining is the process of extracting valid information from large databases. Data mining used in many research areas. Data mining is widely used in agricultural data processing. There are many techniques of Data mining which are developed for decision making (DSS). If proper decision making techniques are applied on data then these data are stored in data bases. These data are used to understand the hidden correlation between crop-pest and disease-weather parameters eg Rapid association rule mining (RARM) is used to find out correlation between various weather and parameters w.r.t pests. There are also precision process model which has been extended to real time DSS.

### 2. Normalize and standardized data

In data mining, KDD process is a two step process- first is pre processing and second is post processing. These processing plays an important role. Various parameters of data are deal in these processes. These parameters have scales which are used in many data mining techniques. These scales help to measure the distance between two objects for example Euclidean distance.

#### Euclidean distance

Euclidean distance is used in data mining for to measures of distance between samples. This helps in faster clustering. In mathematics distance between two points in Euclidean space is Euclidean distance. This distance is defined by square root of sum of squares of distance between two points or samples. These points are connected by a straight line for example there are two points p and q with dimensions as  $p(p_x, p_y)$  and  $q(q_x, q_y)$  define as-

$$\text{Distance}(p,q)=((p_x-q_x)^2+(p_y-q_y)^2)^{1/2}$$

Normalize means in mathematics divide a vector by its magnitude to produce a unit vector and standardization means to transform it to have zero mean (subtraction of mean). Normalization scales in data in short interval like [0,1] and data in standardization are unbounded.

### 3. Wireless sensor network in data mining

Many Wireless sensor networks are made for real-time decision. The main aim of Wireless sensor network is to gathering of knowledge which helps in making outline of real-time decision to understand. Wireless sensor network helps in increasing space and time resolution to crop soil weather environment. Some experiments were done in order to understand the hidden correlation between crop-pest and disease-weather parameters. The reason behind this is the increasing population and most of the population directly depends on agriculture. Some data mining techniques are applied for forecasting the weather conditions for future loss. This also helps to study soil characteristics. But some data mining techniques have not been applied which help in agricultural problems.

Wireless sensor network consists of nodes which are the small electronic components. These nodes gather the information from surrounding. This information may be in the form of Temperature, Age of the Crop, Leaf Spot, Wind Speed, Leaf Wetness, Rain Fall, Rain Days, etc.

#### 4. Data mining steps

1. First step of data mining is to understand the domain.
2. Then extract the information from that domain.
3. After extracting the information transform it onto data ware house system.
4. Then apply the DM Methods on transformed dataset in order to extract data patterns.
5. When patterns are obtained analyze that data.
6. At the end, represent that data in format in knowledge form.

#### 5. Data mining parameters

1. Clusters are used in data mining algorithm for data analysis.
2. Feature extraction – for to extract the features from datasets.
3. Association- to find the rules between datasets.
4. Discover the patterns in data for prediction.
5. Decision support system for path analysis.
6. Probabilities network for future state by the use of set of rules.

#### 6. Machine learning

Data mining helps is categorization and analyzing data. Machine learning is a branch of artificial intelligence. It is the study of system that can learn from the system. Arthur Samuel (1959) defined machine learning. According to him it is a field of study to learn without explicitly programmed. Machine learning is a learning process[7]. It has the ability to learn from its environment and improve its performance. There are three type of machine learning-

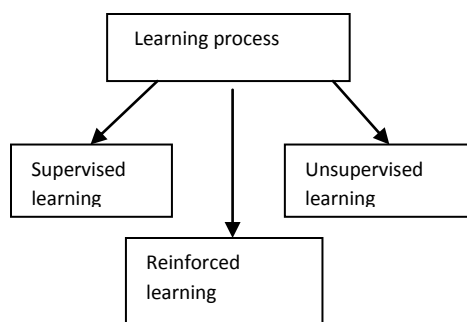


Fig 6.1 types of learning process

- 1) Supervised learning- in this learning teacher is involved. Teacher is having knowledge of the environment. This knowledge is used as set of input and output; here priori knowledge is present which help in matching output patterns. It is self-supervised learning.
- 2) Unsupervised learning-in this learning no teacher is present. So no priori knowledge is present which can check output is coming is correct or not and also called self-organized [7].

- 3) Reinforced learning-it is the intermediate of both supervised and unsupervised learning. Here teacher is present which helps only output is correct or incorrect. This helps for future actions taken.

## 7. Research Design

Research Design includes following steps –

- Accept soyabean dataset then Perform parameter adaptive analysis for attribute selection.
- When analysis is done then perform HMM model on them.
- After then apply fuzzy weight.
- Apply training and testing set on dataset for to divide.
- For parameteric adaptation generate the fuzzy weights
- Neural network is perform for classification
- After then Perform recognition is done
- Last identify the Error Rate and Recognition Rate

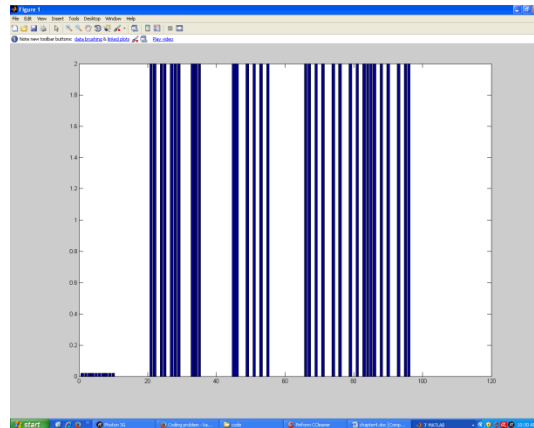
## 8. Experimentation

- To process in quantized form, it is required to convert this dataset in numerical form i.e data transformation
- Statistics analysis applied on each data form so that the data statistics will be processed.

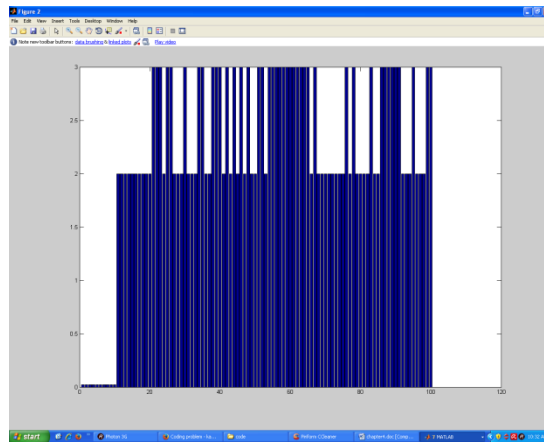
## 9. Result

Data value analysis on three selected features-

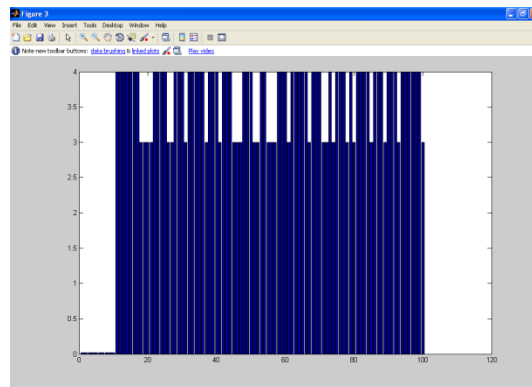
- Plantstand
- Severity
- Crop history



**Figure 9.1:** Data Value Analysis (Plantstand)



**Figure 9.2:** Data Value Analysis (Severity)



**Figure 9.3:** Data Value Analysis (Crop History)

## 10. Conclusion

In this present work, prediction of agricultural disease is done by defining neural and HMM adaptive method. It involved three main stages-On datasets attributes the weight identification is done using fuzzy approach. This is done for disease identification. In second stage of this model, HMM is defined to identify the relative weight identification. This stage is here defined to generate the associated weights. Then neural network is applied to identify the disease.

## References

- [1] Li Zhanfeng, " An Unattended Detection Method of Soil Respiration", 978-1-4244-7161-4/10©2010 IEEE
- [2] Metin Akay, " AUTOMATED NONINVASIVE DETECTION OF CORONARY ARTERY DISEASE USING WAVELET-BASED NEURAL NETWORKS", 0-7803-2050-6@1994 IEEE
- [3] Xavier SIMON, " PLOT ANALYSIS FOR CROP IDENTIFICATION AND DISEASE DETECTION", CH2971-0/91/0000-1927@ 1991 IEEE
- [4] Takashi Kido, " Haplotype Pattern Mining & Classification for detecting disease associated Site".
- [5] A. K. Tripathy, " Data Mining and Wireless Sensor Network for Groundnut Pest/Disease Precision Protection".

- [6] K.Srinivas," Mining Association Rules from Large Datasets towards Disease Prediction", 2012 International Conference on Information and Computer Networks (ICICN 2012) IPCSIT.
- [7] Ting Li," Prediction and Early Warning Method for Flea Beetle Based on Semi-supervised Learning Algorithm", Fourth International Conference on Natural Computation 978-0-7695-3304-9/08 © 2008 IEEE
- [8] Dr. D. Raghu," Probability based Heart Disease Prediction using Data Mining Techniques", IJCST ISSN: 0976-8491 (Online) | ISSN: 2229-4333(Print).
- [9] Anshuk.a Srivastava," Development of a Sensor for Automatic Detection of Downey Mildew Disease", Proceedings of the 2004 International Conference on Intelligent Mechatronics and Automation 0-7803-8748-1/04/@2004 IEEE
- [10] Zhang Lei," Characteristic of lead content in soil and crop in Zhaoyuan gold mine area", 2012 International Conference on Biomedical Engineering and Biotechnology 978-0-7695-4706-0/12 © 2012 IEEE.
- [11] I.V.Bragin," PASSIVE MICROWAVE COMPLEX OF REMOTE SENSING IFOR THE DETECTION OF OBJECTS IN THE SOIL''.
- [12] Mrs.G.Subbalakshmi," Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), ISSN 0976-5166 Vol. 2 No. 2 Apr-May 2011 170-174.
- [13] Abhishek D. Raut and R. D. Raut, "Optimal Classification of Soyabean Crop Diseases using Neural Network Approach", IOSR Journal of Electrical and Electronics Engineering, PP 37-40, 2014.