Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 03, March -2018

CROP YIELD PREDICTION IN AEROPONICS FARMING USING DATA MINING TECHNIQUES

[1]Kiruthika.S, [2]Lakshna.L, [3]Menaka.P,
[4]Dr.R.Madhumadhi, [5]Mr.S.Suresh Kumar

[1]*Sri Ramakrishna Engineering College.*
[2,3,4,5]*VIT University*

ABSTRACT:- *In soilless farming like Aeroponics, the temperature is one of the major affecting factors in producing high yield of crops. The varying seasonal temperature has a high chances of altering the inside room temperature. This paper focuses on predicting the yield of crop in the current year using the historical dataset and in addition, it also focuses on predicting if the yield will go through a downfall with the change in the room temperature. This prediction aids the farmer to get aware of the decline in the yield and take corrective measures in setting the appropriate room temperature to overcome the decline in the yield of crops.*

Keywords: *soilless farming, dataset, yield downfall, data mining.*

1.INRODUCTION

Aeroponics is an indoor gardening practice in which plants are grown and nourished by suspending their root structures in air and regularly spraying them with a nutrient and water solution. Soil is not used for aeroponics, because the plants can thrive when their roots are constantly or periodically exposed to a nutrient-rich mist.
Data mining is an interdisciplinary field that mines into a large volume of data to discover various patterns to extract the desired information and transform it into a comprehensive structure for further use. Yield prediction is an important agricultural problem.
We have conducted an experiment on the tomato crops predicting its yield in the current year by utilizing the historical analysis of the yield. It is known that the plants in aeroponics farming grow under a fixed temperature in a closed environment. With the varying seasons, the air temperature outside the closed environment will also vary which in turn will have an immense impact on the ambient room temperature. This alteration in the room temperature has high chances of resulting in the decline of the crop yield .The goal of our work is to predict if there will be a downfall in the yield with the changing room temperature and to let the farmer know pre-handed to decide on the corrective measures that are to be taken in order to set an appropriate room temperature to overcome the problem of downfall in the yield of the tomato crops.

2.LITERATURE SURVEY

From the research article [3], the researcher expresses that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages.
There are several applications of Data Mining techniques in the field of agriculture. The researchers implemented [4] K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbor is applied [8] for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed[7] using Support Vector Machine
Nutrients descriptions were proposed[1] by the researcher for classifying the soils that is suitable for Aeroponics farming.

3. PROPOSED SYSTEM

This work incorporates two modules. Firstly, predicting the yield of the tomato crop using the historical analysis of the yield. Data mining algorithms like k-Means has been used for the clustering process. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. This algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

The following figure shows the clustering based upon the fixed inside temperature and the seasonal varying temperature.

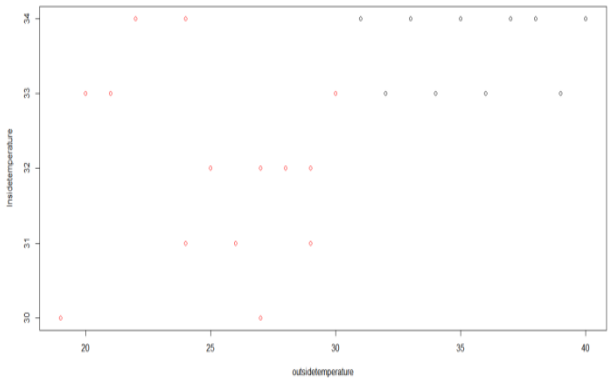

Fig 1: The clustering process using K-Means algorithm.

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for regression and classifiaction. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression.

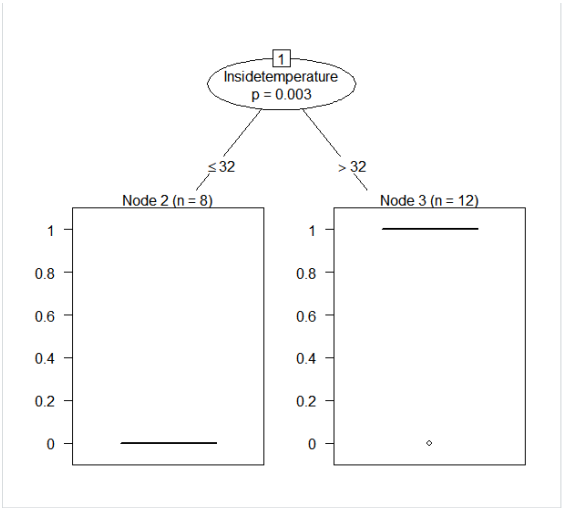

Fig 2: The classification process using K-Nearest Neighbor Algorithm

For prediction, Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. The population regression line for p explanatory variables x_1, x_2, ... , x_p is defined to be $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values b_0, b_1,....., b_p estimate the parameters β_0, β_1....., β_p of the population regression line.

Since the observed values for y vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_p x_p$. The "RESIDUAL" term represents the deviations of the observed values y from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ε.

Formally, the model for multiple linear regression given for n, is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + .. \beta_p x_{ip} + \varepsilon_i$ for $i = 1,2, \ldots n$.

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-squares estimates b_0, b_1, ... b_p are usually computed by statistical software.

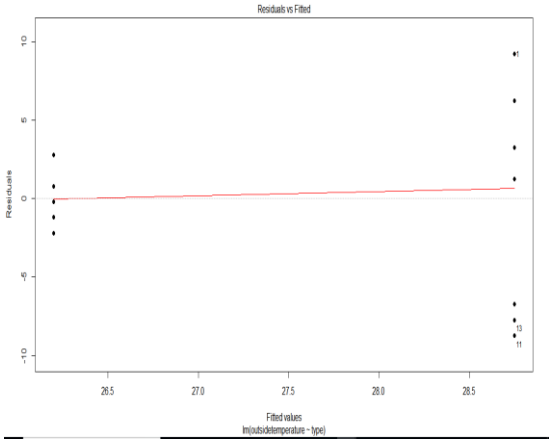

Fig 3: Predicting crop status using Multiple Linear Regression Algorithm

Secondly, Predicting the crop yield based on the previous year dataset using time series modeling for regression. Time series regression is a statistical method for predicting a future response based on the response history (known as autoregressive dynamics) and the transfer of dynamics from relevant predictors. Time series regression can help you understand and predict the behavior of dynamic systems from experimental or observational data. Time series regression is commonly used for modeling and forecasting of economic, financial, and biological systems.

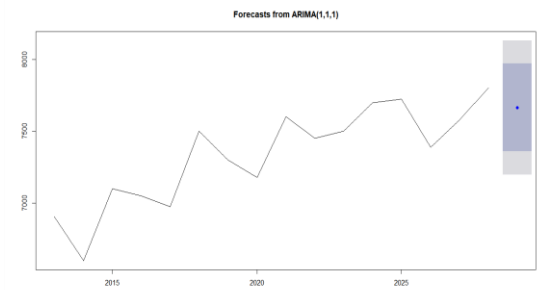

Fig 4: Prediction process using Time series Modeling for crop yield prediction.

4. CONCLUSION

Agriculture is the most important application area particularly in all developing countries like India. Using the computer algorithms in agriculture can change the scenario of decision making and the farmer can produce yield in a better way. In this paper certain data mining algorithms have been adopted to cluster (k-means), classify (k-NN),establish relationship between two or more variable(Multiple Linear Regression) and to predict the future response based on time(Time series regression).The paper's main focus is to help the farmer to solve the problem of decline in the yield of aeroponics to overcome the downfall. This paper integrates the work of various authors and it is useful for the researches to get information of the current scenario of data mining techniques and application in the context of agriculture field.

5.REFERENCES

1. Andrzej, Tomasz, "The effect of nutrient solutions on yield and macronutrient status of greenhouse tomato grown in aeroponic and Rockwool culture with or without recirculation of nutrient solution", Acta Scientiarum Polonorum,2014.
2. G R Batts, "Effects Of CO2 And Temperature on Growth and Yield of Crops of Winter Wheat over Four Seasons", European Journal of Agronomy, vol. 7, 1997, pages : 43-52.
3. G Rub, "Data Mining of Agricultural Yield Data : A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages : 24-37.

4. Jorquera H, Perez R, Cipriano A, Acuna G, "Short Term Forecasting of Air Pollution Episodes", In: Zannetti P (eds) Environmental modeling , WIT Press, UK, 2001.
5. Lemma, Abede, Gebremedhin, "Determination of Nutrient Solutions for Potato Seed Production under Aeroponics Production System", Open Agriculture, 2017
6. Mbiyu, Muthoni, Kabira, "Use of aeroponic technique for potato minitubers production in Kenya", Journal of Horticulture and Forestry, 2012.
7. Tripathi S, Srinivas V V, Nanjundiah R S "Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach", J Hydrol, 2006, pages : 621-640.
8. Tellaeche A, X P Burgos Artizzu, G Pajares, A Ribeiro, "A Vision-Based Classifier for Weeds Detection in Precision Agriculture through the Bayesian and Fuzzy K-Means Paradigms", Adv.Soft. Comp., 2008, pages : 72-79.

