# IOT Based Smart Agriculture System and Crop Recommendation using ML

Raghavendhar T.V[1], Joslin Joshy[2], Bhaskar A[3]
*Department of CSE, SRM Institute of Science and Technology, Tamilnadu, India*

***Abstract-*** The agricultural industry has undergone major developments over the last century. Today's farming involves IOT systems which aids in the enhancement of agriculture yield. Sensors devices detect soil moisture, pressure and alert farmers on their mobile device connected to a GSM module when there is a variation in pressure or soil water levels. Data from these sensors are sent to a cloud storage. Data ingestion to Hive database promotes efficient storage of the data. Machine learning algorithms help to improve the systems in crop recommendation based on the soil type. The system helps farmers to increase productivity and promotes efficient farming.

***Keywords-*** Sensors, Hive, Machine learning, Crop recommendation, GSM module, Arduino

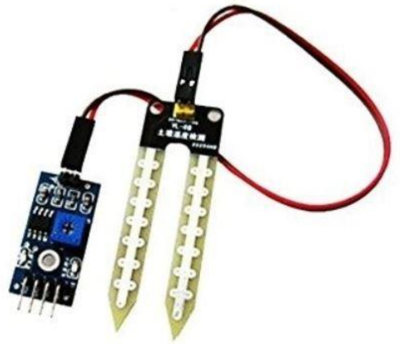
## I. INTRODUCTION

Internet of things has the ability to transform the world and has the capability to atomize any work in an efficient way. Thus, the use of IOT in the field of smart farming has the ability to increase productivity and it is performed with the help of automation. There has been various works carried out in the field of smart farming [1] the crops are monitored by the devices such as Arduino, breadboard, Wi-Fi module, soil moisture sensor and temperature sensor.Live data feed is obtained from thingsspeak.com. It help the farmers to receive data regarding the soil moisture and temperature, thus helps in continuous monitoring of the agricultural fields. In this we are going to improve the system by using GSM module. The soil moisture sensor, pressure sensor, water level detector are connected to GSM module through which messages are sent to the farmers. Huge dataset can be handled efficiently using Big Data [2] Big Data technology with Hadoop for distributed computing transfers the relational database table through the tool kit called as Sqoop and is stored in Hbase. To access the values of Hbase, hive query processing is done which is used for the further analysis. This system stores the data in cloud through NodeMCU. The whole data is send to hive using Python. With the help of cloud computing the data is accessed virtually. [3] The goal is to predict crop yield based on the various needs necessary for the development of the crop. The yield analysis and crop prediction is performed using the data. Extraction technique based on the association rule for a particular region. ie District of TamilNadu, India. There are several studies that are conducted in the field of Big Data, [4] the study of demographics in Big Data Technology is related to the statistical measure that is the study of the human population in a given period of time. It focuses mainly on birth, death rate, literacy rate and marriages. In the computer age, the field of social sciences is merging with the field of computing to exploit the ability to collect and analyze data on a large scale. This study deals with Big Data technology with Hadoop, which Map uses to reduce the work of frames for distributed processing. In this case, the table values in the relational database are transferred via the toolkit called Sqoop and stored in Hbase. To access Hbase values, Hive query processing is used, which is used for the next analysis. Since then, the hive has some high latency restrictions, however, the Hbase lacks analytical capabilities. This technology will produce the best possible solution for analyzing data. [5] Prediction of yield according to the various requirements necessary for the development of crops. This paper presents a brief analysis of crop yield and prediction using a data mining technique based on the association rule for the selected region. That is, the Tamil Nadu district in India. The experimental result shows that the proposed work efficiently provides crop yield. For this study, all the required data for the crop prediction are collected which are used in testing and training phases. The training phases consists of preprocessing, clustering Data, Applying Associate rule mining and generating rules. In the testing phase the user gives the requirement based on which the predictions are carried out.This work suggests crop only for Tamil Nadu district which has been improved by our system by providing diverse suggestions for all the states of India. The experimental result shows the accuracy of crop prediction and recommendation. Thereby, helping the farmers to increase the yield. The recommendation system compares two machine learning algorithms, KNN and SVM to find the most accurate algorithm which recommends crops based on soil type. The system also generates daily reports to farmers regarding Air pressure, Soil moisture, Temperature and Humidity.

## II. COMPONENT DESCRIPTION

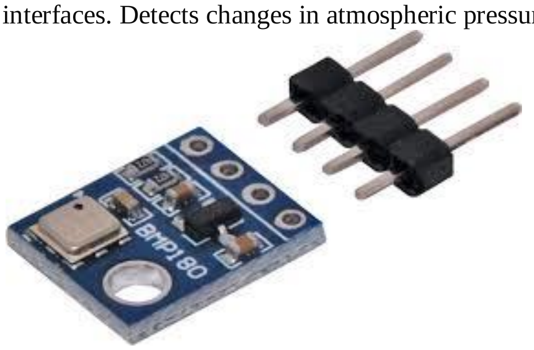
### *A. Soil Moisture Sensor*

The soil moisture sensor has two probes that are used to measure the volume of volumetric water. The probes allow the current to pass through the ground and use the resistance value to measure the moisture value. The ground will conduct more electricity when there is more water, which means there will be less resistance. Therefore, the humidity level will be high. The dry soil badly conducts electricity, so that when there is less water, the ground will conduct less electricity, which means that there will be more resistance. Therefore, the humidity level will be low.



*B. Air Pressure Sensor*

The BMP180 is a high-precision pressure sensor with low power consumption. It can be used on smartphones, tablets and sports devices. It is followed by the BMP085 and brings many improvements, such as smaller size and expansion of digital interfaces. Detects changes in atmospheric pressure.

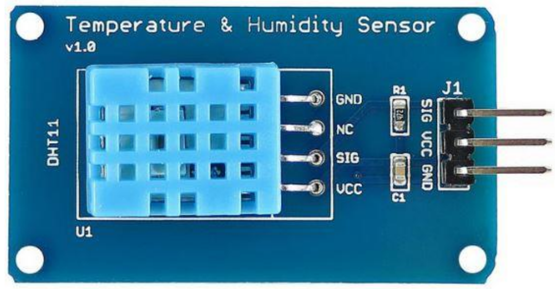



*C. Water Level Detector*

The Arduino water level detection sensor has a DC3-5V operating voltage and the operating current is less than 20 mA. The sensor is of the analog type which produces analogue output signals based on water pressure. Conversions of the water signal on an analog signal are easy and the analogue output values can be read directly using the Arduino development board, which helps to achieve the level alarm effect.

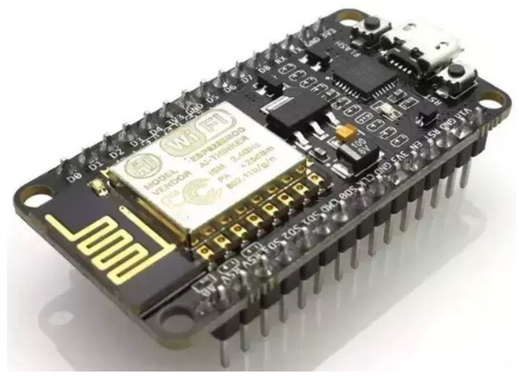
*D. Temperature and Humidity Sensor*

The DHT11 temperature and humidity sensor has a calibrated digital signal output together with the capacity of the temperature and humidity sensor. It is integrated with a high-performance 8-bit microcontroller that uses a capacitive humidity sensor and a thermistor to measure the surrounding air.



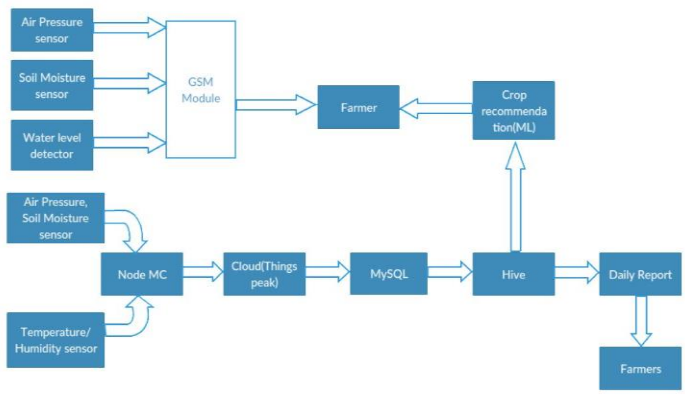

*E. Node MCU*

NodeMCU is a WiFi module based on ESP8266 -12E. It is a highly integrated WiFi system on a chip designed to provide complete Internet connectivity in a small package. It can be programmed directly via the USB port using Arduino IDE. NodeMCU is the WiFi module equivalent to the Ethernet module, as it combines the features of the WiFi access point, the station and the microcontroller. It can be used as a host of a web server or can be connected to the Internet to restore or load data.



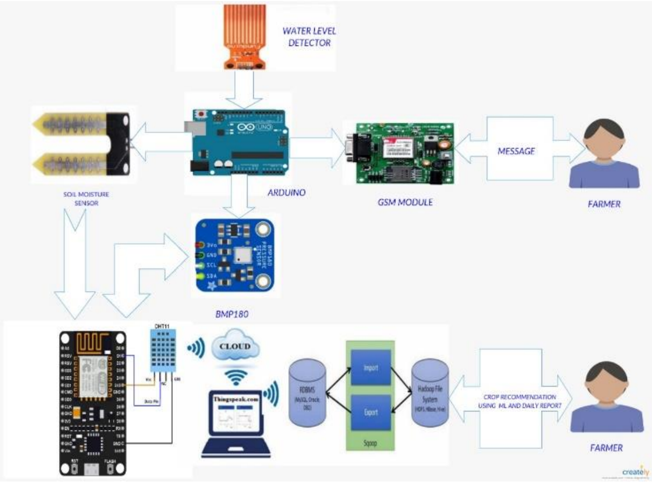
## III. WORKING OF THE SYSTEM

*A. Proposed structure of the system*




The structure represents the smart agriculture system for crop recommendation

*B. Detailed Description of the Operation*




*1) Step 1: Air pressure sensing*

The BMP180 barometric pressure sensor can be used to predict the weather. It's an extremely sensitive sensor. Air pressure values are detected every second. The sensor sends a cyclone alert to the farmers when there is a rise in air pressure levels. It is done with the help of a GSM module which is connected to the sensor.

*2) Step 2: Soil Moisture sensing*

Soil moisture is measured based on the water content present in the soil. It is measured using a soil moisture sensor consisting of two conductive probes, capable of measuring the moisture content in the soil as a function of the resistance variation between the conductive probes. The sensor sends notifications to farmers based on the soil moisture content. This helps farmers to regulate the amount of water supplied to the soil.

*3) Step 3: Water Level Detection*

The water level detector is capable of detecting the level of water thereby alerting the farmer in case of high water levels which may result in floods. Water levels are detected every second and the farmers are alerted when excessive water levels or extremely low water levels are detected, which may destroy plant growth. A GSM module is connected to the sensor which sends notifications to the farmers.

*4) Step 4: GSM Module*

The GSM module is interfaced with Arduino. It is primarily used to send and receive messages. This module is connected to the air pressure sensor, soil moisture sensor and water level detector. The sensory values from the sensors are analyzed and notifications are sent to the farmers on their smart device. Thereby helping the farmers to keep track of the farming conditions.

*5) Step 5: Temperature and Humidity Sensing*

DHT11 temperature and humidity sensors work by detecting changes that alter electrical currents or temperature in the air. These values help to detect the appropriate crops that can be produced in that region based on their temperature and humidity conditions. This sensor is connected to the MCU node for further processes.

*6) Step 4: NodeMCU*

To communicate and transfer data to the cloud in real time, there is another chip called NodeMCU that can communicate directly with live servers. NodeMCU integrates the firmware with integrated Wi-Fi connectivity so that direct interaction with the network can be performed without physical jumper cables. This is an Arduino-like plate that is used to interact with the cloud in real time, so that signals or live data can be recorded or archived for predictive analysis. It can be programmed via the USB port using Arduino IDE. The air and humidity pressure sensor together with the temperature / humidity sensor produce data in real time every second. This data is received via the MCU node. This data is sent to the cloud storage, which is ThingSpeak.

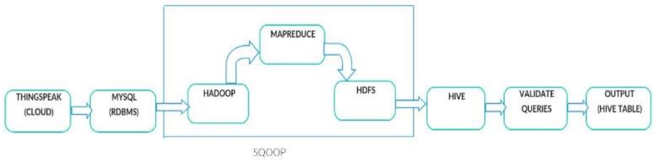
*7) Step 5: Cloud storage*

Data stored in cloud can be easily accessed. NodeMCU sends live data to ThingSpeak, which can be analyzed and visualized as graphs or chats. It also helps in generating daily reports by scheduling the operation to run at a particular time. This helps the farmers to visually understand the changes in the soil and climatic conditions using built in plotting functions.

*8) Step 6: MySQL*

MySQL is a widely used Relational Database Management System (RDBMS). MySQL server can manage many databases at the same time. But when it comes to large dataset, MySQL is not as efficient as Hive database. MySQL stores data by horizontally partitioning the dataset. RDBMS is easily accessible by cloud and live data can be constantly updated into the SQL database.

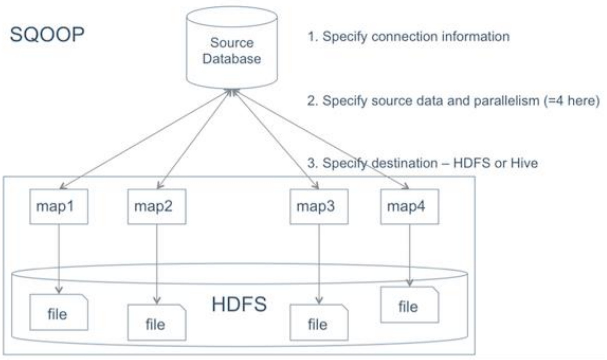
*9) Step 7: Hive database*




The above flow diagram represents the auto ingestion of the cloud data that is stored in MySQL to Hive. Hive is an efficient database that is used to store huge data efficiently. It takes redundant copies of the database thus avoiding loss of data. This is performed using Sqoop tools.

Sqoop**:**

Sqoop is a tool that allows you to import and export data in bulk from a database. You can use Sqoop to import data to HDFS or directly to Hive. Sqoop is an Apache tool that allows you to import or export data in bulk from a database such as MySQL. In this system Sqoop jobs are used in order to parallel update the database from MySQL to Hive. Sqoop uses map reduce function which internally transfers data in Hadoop to HDFS. The diagram represents how Sqoop is used to move data into Hive.




*10) Step 8: Comparative analysis of algorithm to determine accuracy*

The *Graph:1* represents the accuracy between k-means and SVM algorithms. The accuracy percentage of the algorithms are measured using precision metrics in order to determine the most accurate algorithm thereby helping the farmers determine the crops that are to be planted for the respective soil type.

***Algorithm 1: K-means***

K-means clustering is a type of non-supervised learning, used when you have data without tags (ie data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works in an iterative way to assign each data point to one of the K groups based on the characteristics provided. Data points are grouped according to the similarity of characteristics. The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data
2. Training data labels (each data point I is assigned to a single group)

Instead of defining groups before examining data, grouping allows you to search and analyze groups that have been trained organically. The "Choice of K" section below describes how to determine the number of groups. Each centroid of a cluster is a collection of characteristic values that define the resulting groups. The examination of the weights of the centroid characteristics can be used to qualitatively interpret which type of group represents each group.

The data set is given as the input. The soil type is specified, according to which the Dataset is analyzed to find the right crop. The temperature, humidity, soil moisture and air pressure values are fetched from the Dataset, on which the k-means clustering is performed. K-means clustering aims to divide n observations into k clusters in which each observation belongs to the cluster with the nearest average. This involves a data set partition in different clusters. The k difference between each cluster is found. The difference values represents a specific crop according to the soil type. Similarly, the soil moisture, atmospheric pressure, humidity and temperature are determined. The k value for each parameter is stored in the database, based on which the appropriate crop is recommended.

***Algorithm 2: Support Vector Machine***

"Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification and regression challenges. However, it is mainly used in classification problems. In this algorithm, we trace each element of data as a point in the n-dimensional space (where n is the number of entities it has) with the value of each entity as the value of a particular coordinate. So, we perform the classification by finding the hypothalamus that differentiates the two classes very well (see the snapshot below). Support vectors are simply the coordinates of individual observation. The Support Vector Machine is an edge that best segregates the two classes. The SVM model is a representation of the examples as points in space, assigned so that the examples of the separate categories are divided by the widest possible free space.In addition to performing linear classification, SVMs can perform a nonlinear classification efficiently, implicitly mapping their voices into high-dimensional feature spaces.

- It works very well with a clear margin of separation.
- It is effective in high-dimensional spaces.
- It is effective in cases where the number of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision-making function (called support vectors), so it is also efficient in terms of memory.

*11) Step 9: Crop recommendation*

According to soil type, the crop recommendation is done by considering all the parameters such as air pressure, soil moisture, temperature and humidity. The system is capable of recommending appropriate crops based on the predictions done using k-means and SVM algorithms. Recommendations are done for a particular soil types such as black soil, red soil, dessert soil, etc.

*12) Step 10: Daily reports*

Daily reports are generated from ThingSpeak to the farmers through mail. Reports on Soil moisture, Air pressure, Temperature and humidity are generated as bar graphs as shown in *Graph: 2 to 5*. These graphs help to keep track of daily events.

## IV. SNAPSHOTS OF SYSTEM

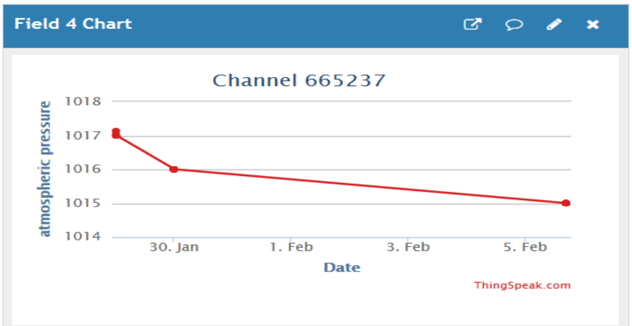



Fig.1: Air pressure graph

Figure.1 shows the atmospheric pressure value of four alternative days. This graph is an analysis on the data present in cloud. i.e. ThingSpeak which is obtained from the sensor through NodeMCU. Variations indicate the changes in air pressure for the respective days. It helps in detecting cyclone which can be caused by low air pressure.

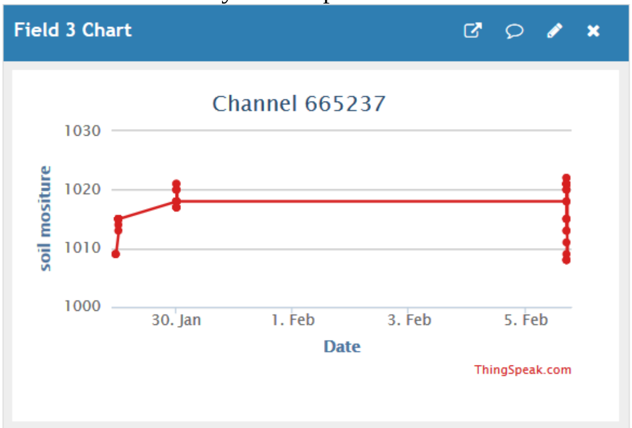



Fig.2: Soil Moisture graph

Figure.2 shows the soil moisture levels of four alternative days. This graph is an analysis on the data present in cloud. i.e. ThingSpeak which is obtained from the sensor through NodeMCU. Variations indicate the changes in soil moisture levels for the respective days. It helps in detecting floods when there is a rise in the water level above a certain safe limit. And also alerts the farmers when there the soil moisture level is low.

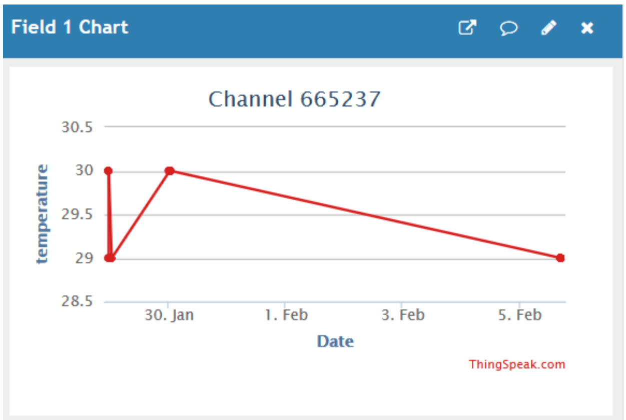



Fig.3: Temperature graph

Figure.3 shows the temperature values of four alternative days. This graph is an analysis on the data present in cloud. i.e. ThingSpeak which is obtained from the sensor through NodeMCU. Variations indicate the changes in temperature for the respective days which helps in analyzing appropriate crops for a region based on weather conditions.

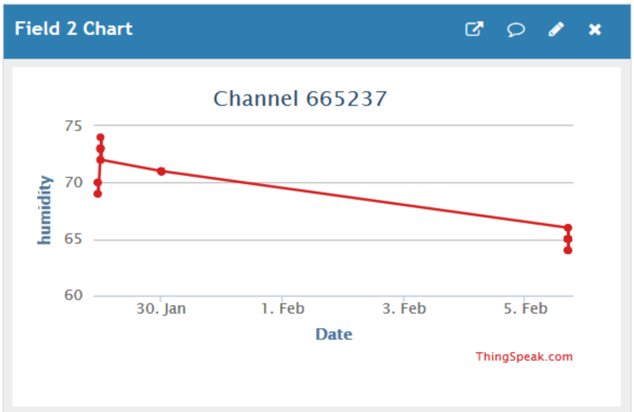



Fig.4: Humidity graph

Figure.4 shows the humidity values of four alternative days. This graph is an analysis on the data present in cloud. i.e. ThingSpeak which is obtained from the sensor through NodeMCU. Variations indicate the changes in humidity for the respective days which helps in analyzing appropriate crops for a region based on weather conditions.

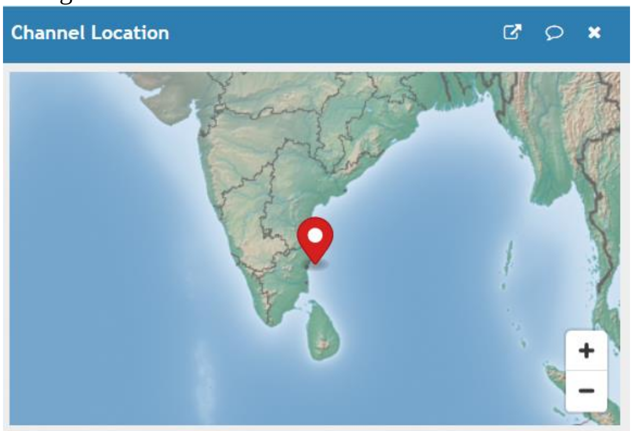



Fig.5: Location

Figure.5 shows the location for which the air pressure, soil moisture, temperate and humidity values and analyzed. This helps to view and analyze a region efficiently.

*MySQL:*

```
mysql> select * from test1;
```

| timething | id | temp | hum | moist | atm | time |
|---|---|---|---|---|---|---|
| 2019-01-15 21:58:01 | 1 | 28.00 | 75.00 | 1020 | 1014.71 | 2019-01-16 03:31:49 |
| 2019-01-15 21:58:24 | 2 | 28.00 | 75.00 | 1020 | 1014.70 | 2019-01-16 03:31:49 |
| 2019-01-15 21:58:48 | 3 | 28.00 | 75.00 | 1024 | 1014.67 | 2019-01-16 03:31:49 |
| 2019-01-15 21:59:11 | 4 | 28.00 | 75.00 | 1021 | 1014.67 | 2019-01-16 03:31:49 |
| 2019-01-15 21:59:35 | 5 | 28.00 | 75.00 | 1015 | 1014.64 | 2019-01-16 03:31:49 |
| 2019-01-15 21:59:59 | 6 | 28.00 | 75.00 | 1023 | 1014.76 | 2019-01-16 03:31:49 |
| 2019-01-15 22:00:22 | 7 | 28.00 | 75.00 | 1021 | 1014.68 | 2019-01-16 03:31:49 |
| 2019-01-15 22:00:46 | 8 | 28.00 | 75.00 | 1021 | 1014.68 | 2019-01-16 03:31:49 |

Table 1. Data stored in MySQL

The data from cloud is fetched into MySQL as a comma separated value format or as a text file. Table 1 shows the values stored in test1. It consists of parameters such as temperature, humidity, soil moisture and air pressure values obtained at different time intervals.

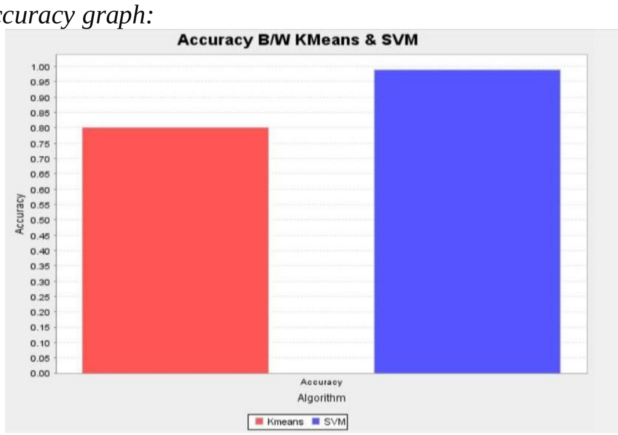
*HIVE:*

```
hive> select * from test1;
OK
2019-01-15 21:58:01     1       28.00   75.00   1020    1014.71 2019-01-15 14:01:49.0
2019-01-15 21:58:24     2       28.00   75.00   1020    1014.70 2019-01-15 14:01:49.0
2019-01-15 21:58:48     3       28.00   75.00   1024    1014.67 2019-01-15 14:01:49.0
2019-01-15 21:59:11     4       28.00   75.00   1021    1014.67 2019-01-15 14:01:49.0
2019-01-15 21:59:35     5       28.00   75.00   1015    1014.64 2019-01-15 14:01:49.0
2019-01-15 21:59:59     6       28.00   75.00   1023    1014.76 2019-01-15 14:01:49.0
2019-01-15 22:00:22     7       28.00   75.00   1021    1014.68 2019-01-15 14:01:49.0
2019-01-15 22:00:46     8       28.00   75.00   1021    1014.68 2019-01-15 14:01:49.0
2019-01-15 22:03:23     9       28.00   77.00   1020    1014.74 2019-01-15 14:03:51.0
2019-01-15 22:03:47     10      28.00   75.00   1021    1014.67 2019-01-15 14:03:51.0
2019-01-15 22:26:12     11      28.00   75.00   1020    1014.63 2019-01-15 14:26:23.0
2019-01-15 22:26:35     12      27.00   74.00   1020    1014.63 2019-01-15 14:26:45.0
```

Table 2. Data stored in Hive

The data from MySQL is fetched into Hive with the help of sqoop jobs. Table 1 shows the values stored in Hive database. It consists of the same parameters that are fetched from MySQL such as temperature, humidity, soil moisture and air pressure values obtained at different time intervals. The data in MySQL is parallel updated into Hive through auto ingestion of data.

*Accuracy graph:*

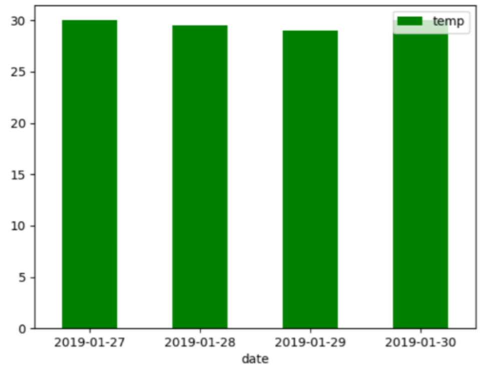



*Graph: 1*

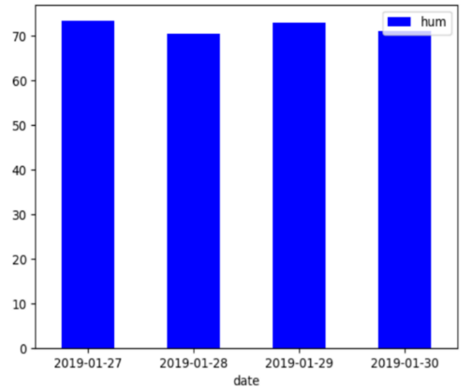
From the comparative analysis of Support Vector Machine (SVM) and K-Means algorithm. The accuracy is measured using performance metrics. It was found that SVM algorithm produces highly accurate results when compared to k-means.

*Daily Reports to farmers:*

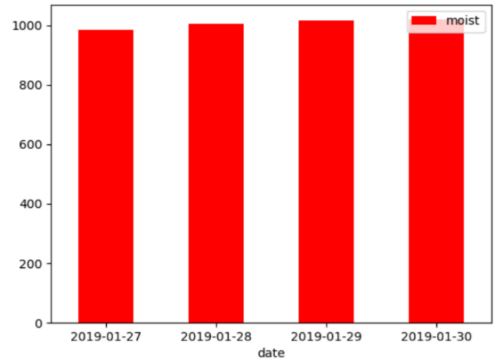
Reports are send to the farmers through mail. It contains reports on Temperature, Humidity, Soil moisture and Air pressure which is generated every day. This helps the farmers to keep track of the ongoing factors that affect the growth of their plants and in efficient farming. The reports are generated as shown in graphs 2 to 5.



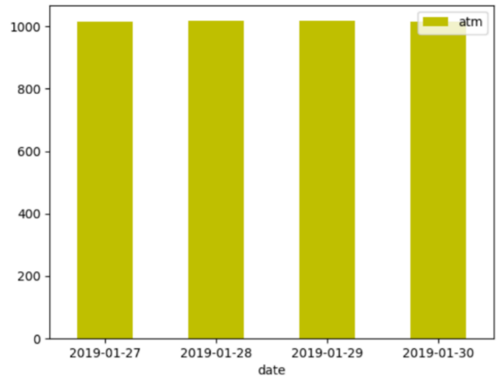

*Graph 2. Temperature report of 4 days*




*Graph 3. Humidity report of 4 days*

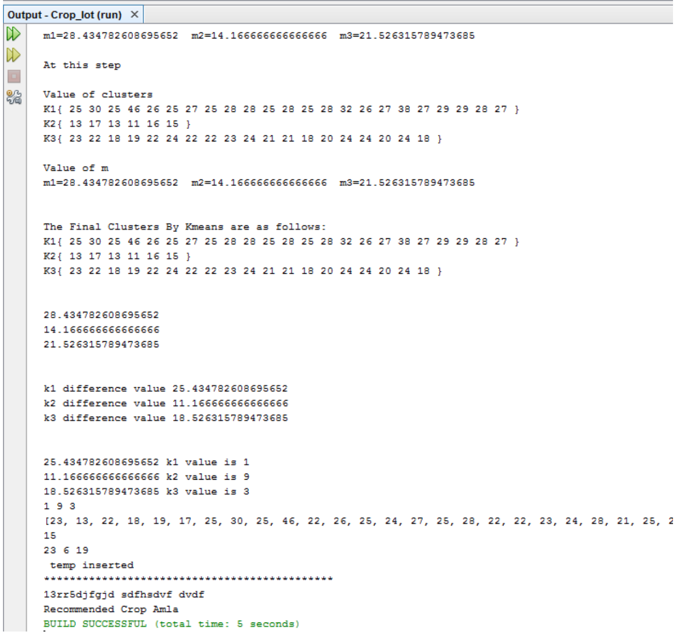



*Graph 4. Soil moisture report of 4 days*



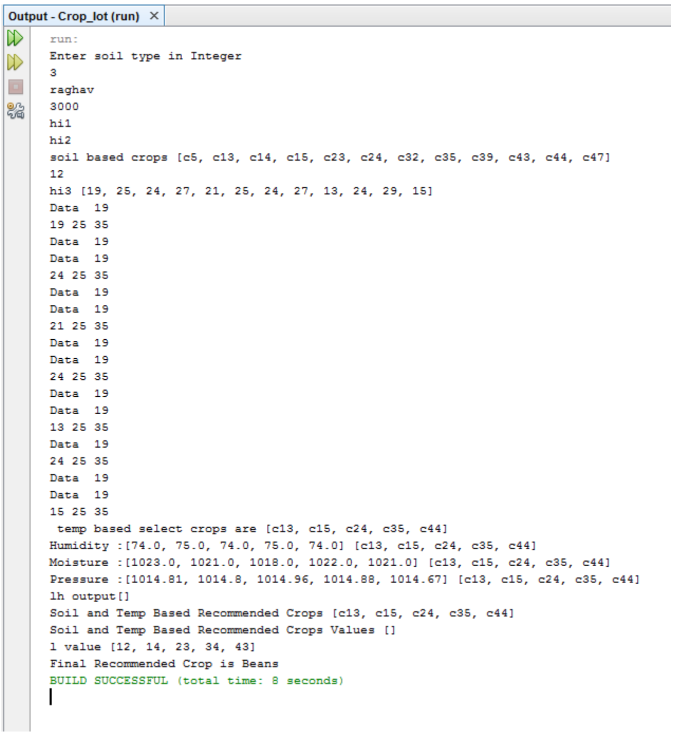

*Graph 5. Air pressure report of 4 days*

*Crop Recommendation:*




*Crop recommendation of K-means*

The output screen shows the crop recommended using k-means algorithm. It is done using Java. Means value is taken as m, using which k clusters are formed. The mean value of these clusters determine the crop according to its soil type. The crop recommended by k-means for a certain soil type is amla.

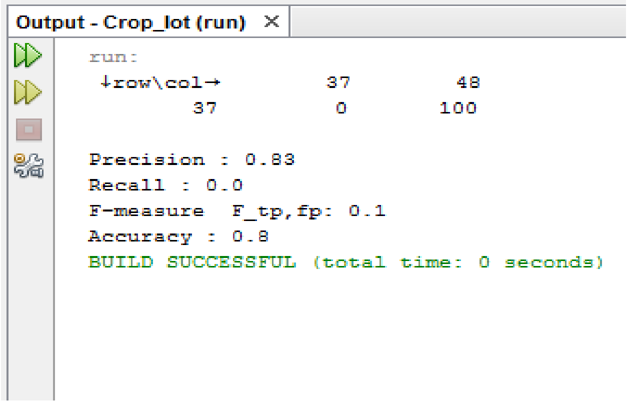



*Crop recommendation of SVM*

The output screen shows the crop recommended using Support Vector Machine algorithm. Here, the soil type is given as the input. The relevant data are fetched and classified according to the required soil type. The crop is determined according to the relevant air pressure, Soil moisture, Temperature and Humidity. The crop recommended by SVM is beans. This is found to be more accurate using precision algorithm.

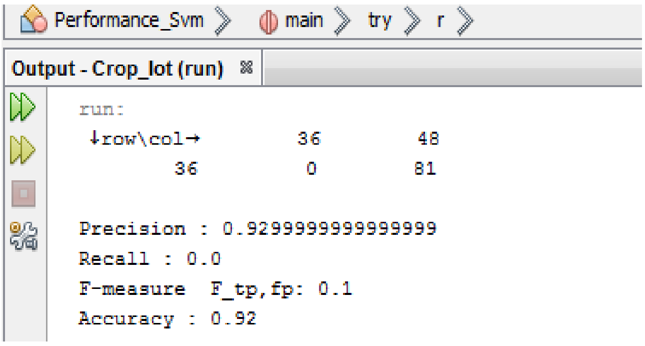
*Prediction accuracy:*

K-means algorithm has an accuracy percentage of 0.8.




*Accuracy of k-means*

SVM algorithm has an accuracy percentage of 0.92. Thus is more accurate.




*Accuracy of SVM*

## IV. CONCLUSIONS

This work mainly focuses on efficient farming by recommending the right crops to the farmers and alerting them in case of emergency. Comparative analysis of the machine learning algorithms and their performance is done in order to find the most accurate algorithm, using which the crop are recommended. On analysing the performance, SVM algorithm has greater accuracy which effectively suggests crops according to the soil type and other parameters. The system has been improvised in terms of storage, accuracy and predictions. This work can be further improved by classifying permanent and non-permanent crops which helps the farmers to grow multiple crops in the span of a year and increase productivity.

## V. REFERENCES

[1]. AnandNayyar, Er. VikramPuri "*IOT based smart sensors Agriculture stick for live temperature and moisture using Arduino, cloud computing & solar technology*"

[2]. Syeda Sana Bukhari, JinHyuckPark,DongRyeol Shin *"Hadoop Based Demography Big Data Management System"* IEEE pp.978-1-5386-5886-5/18, June 2018

[3]. Jiao Guo, Peng-Liag Wei, Jian Liu, Biao Jin, Bao-Feng Su and Zheng-Shu Zhou *"Crop Classification based on Differential characteristics of H/α Scattering parameters for Multitemporal Quad and Dual-polarization SAR Images"* IEEE Trans. Geosci. Remote Sens., pp. 0196-2892, April 2018

[4]. E.Manjula, S.Djoditachoumy.*"A Model for prediction of Crop Yield"*.pp.Vol 6: No, 4, March 2017.

[5]. S.Pudumalar, E.Ramanujam, R.HarineRajashree, C.kavya, T.kiruthika, J.Nisha."*Crop Recommendation System for Precision Agriculture*" (2017). Crop recommendation system for precision agriculture. 2016 Eighth International Conference on Advanced Computing.

[6]. The IEEE website. [Online]. Available: http://www.ieee.org/

[7]. https://thingspeak.com/channels

[8]. Na, S., Xumin, L., & Yong, G. "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm". 2010 Third International Symposium on Intelligent Information Technology and Security Informatics.

[9]. Gandhi, N., Armstrong, L. J., Petkar, O., &Tripathy, A. K."Rice crop yield prediction in India using support vector machines" 2016 13th International Joint Conference on Computer Science and Software Engineering.