UNIVERSITY of RWANDA

COLLEGE OF SCIENCE AND TECHNOLOGY

ACEIoT

Website: www.aceiot.ur.ac.rw
Mail: aceiot@ur.ac.rw

AFRICAN CENTER OF EXCELLENCE IN INTERNET OF THINGS

EMBEDDED COMPUTING SYSTEM

P.O.BOX: 3900 Kigali, Rwanda

MASTERS PROJECT THESIS

"IOT BASED MONITORING SYSTEM OF SOIL PROPERTIES"

A dissertation submitted in partial fulfilment of the requirements for the degree of

MASTERS DEGREE IN INTERNET OF THINGS- EMBEDDED COMPUTING SYSTEM

(MSC in IoT-ECS)

Submitted by

GASANA James Madson (RN: 220014138)

January 2022

DECLARATION

I GASANA James Madson hereby declare that the thesis research entitled “IOT BASED MONITORING SYSTEM OF SOIL PROPERTIES” presented for the purpose of fulfilment of the requirement for the award of Masters in Internet Of Things, Option of Embedded Computing Systems submitted to the University of Rwanda school of ACEIOT is original work under the supervision of Dr. Christine Niyizamwiyitira, Co-supervised by Dr. Pierre Bakunzibake. No part of this thesis has been previously submitted by me or anyone for award of any other degree of any university or institution.

Student name and Registration Number

GASANA James Madson 220014138

Date: 4th January 2022

CERTIFICATION

This is to certify that the thesis work entitled **"**IOT BASED MONITORING SYSTEM OF SOIL PROPERTIES**"** Is a record of the original work done by **GASANA James Madson (220014138)** in partial fulfilment of the requirement for the award masters of science in internet of things at College of Science and Technology, University of Rwanda during the academic year 2020-2021.

This work has been submitted under the guidance of Dr**.** Christine NIYIZAMWIYITIRA and

Dr. Pierre BAKUNZIBAKE

Main supervisor:

Dr. Christine NIYIZAMWIYITIRA

Date and Signature:

Co-supervisor:

Dr. Pierre BAKUNZIBAKE

Date and Signature: **5th January 2022**

Head of master's studies and Trainings: Dr. James RWIGEMA

Date and Signature:

ACKNOWLEDGEMENTS

It is a great pleasure to acknowledge the efforts and contribution of many people whose names may not appear but whose cooperation, friendship and understanding was crucial.

We are fortunate to have worked on this thesis with classmates who are more knowledgeable in database systems who contributed much for deep understanding of it.

We would like to thank entire team that supported in one way or another to come up with this thesis. We would like to thank RAB and Dr.Rutebuka Alex who was contact person for the support during data collection.

Finally would like to acknowledge the support of my supervisors Dr. Niyizamwiyitira Christine and Dr. Bakunzibake Pierre for more inputs and corrections so that we came up with this thesis. We cannot end this without extending the appreciations to UR-ACEIOT staff and administration who provided means in terms of academia and financial support to achieve this thesis.

ABSTRACT

Background: In Rwanda and all around the world agriculture plays an important role for living of human beings for survival and economic growth. Different research shows that as population increases the area of cultivation that is used for habitation and need of food is increasing while farming land is reducing.Smart farming is needed to maintain and secure food availability for this generation and future by using IoT.

Objective: The overall objective of this thesis is to monitor soil properties by providing real time data of a given land to farmers and decision makers like RAB for proper use and productivity of land and predicting the appropriate croptype in that land measured.Monitoring soil properties helps to monitor degradation of land and know when and how much fertilizer needed in that land based to the croptype you want to plant meanwhile system will keep showing the croptype based to whatever fertilizers available.

Methodology: In this thesis,the Descriptive quantitative research approach were used by establishing the relationship between independent variables(NPK,temperature,humidity and PH) and dependent variable(croptype like cereals,legumes,vegetables) where the croptype will depend to the soil properties available in the land.The study provides the way of knowing the current soil properties and appropriate croptype.With this system a farmer can monitor the land nutrients in the given land and can be able to add only needed fertilizers.Again in this system there is model to predict the croptype of a given land based to its soil properties.The data was collected using sensors NPK sensor,PH sensor,temperature sensor,humidity sensor.Data from sensors are sent to microcontroller through direct connection and then sent from microcontroller to the database hosted on the cloud via gprs module and also the realtime data visible on LCD. Data was analysed by using machine learning algorithm known as classification.Among three common types of classification logistic regression,decision tree classifier and knearest neighbor(KNN),we chosed to use decision tree classifier because it provided higher accuracy(99%) than others.

Results: In this thesis,results show that the prototype build made up of sensors,microcontroller with internet enabled is able to collect data and send them to database hosted on cloud and those data are processed to make model then we are able to predict the best fit of croptype based to the soil properties measured.in general the expected output was achieved.The main contribution of this work is that, a farmer can know the status soil properties of land by using

IoT sensors instead of tradition means of using laboratory which is expensive,time consuming and tedious,database hosted on the cloud helps the decision makers to visualize and monitor changing rate of soil properties of a given which can help to mitigate the land use in future and finally we can say using machine learning in prediction of best fitting croptype in any given land is a great contribution.

Keywords: IoT, NPK sensors, soil properties,Database,monitoring and Machine Learning Algorithm, Classification

LIST OF ABBREVIATINS AND ACRONYMS

ACEIOT	: African Center of Excellence in Internet of Things
ECS	: Embedded Computing Systems
FAO	: Food and Agriculture Organization
GPRS	:General Packet Radio Service
GSM	: Global System for Mobile communication
IDE	: Integrated Development Environment
IOT	: Internet of Things
LCD	:Liquid Crystal Display
ML	: Machine Learning
NPK	:Nitrogen Phosphorus Potassium
SIM	: Subscriber Identity Module
UR	: University of Rwanda

LIST OF FIGURES

LIST OF TABLES

Contents

CHAPTER 1.INTRODUCTION

Crop production is the main focus nowadays in the whole world and due to rapid growth of population worldwide. Fertile land to cultivate is not increasing instead it is reducing due to the fact that as people are increasing they need land for shelter and other facilities. With the mentioned above needs there is also issue of land degradation and climate change, therefore all these put pressure on governments and researchers to find solution of how to keep feeding the population, since the farming land remain constant, thus after sometime it loses its natural fertility and due to those chemicals/organic manure is needed to restore soil fertility and these chemicals needs to be monitored and regulated and added adequately. This thesis will contribute to the proper soil properties data storage and monitoring.

1.2 Background and Motivation

The Internet of Things (IoT) has become increasingly attractive to industry and emerges to the public for daily life. Its popularity is largely derived from the impression/concept of widely available connection and easily accessible interaction network. In the IoT world, better decisions can be made automatically without people's involvement or less if any. Nowadays, people can be able to obtain information about anything anytime anywhere through the identification of "Things" regardless of its nature. In this field of IoT, it is anticipated that since network will be capable of collecting and transmitting data. Thereafter, data is automatically analysed and it will bring optimization of the solutions to different circumstances experienced in any field of life (Keyuret al, 2016)

IoT in Agriculture, based on the report of Food and Agriculture Organization of the United Nations(FAO) predicts that the global population will reach 8 billions people by 2025 and 9.6 billion by 2050 (FAO,2009).This means that there is need of modernization in agriculture to be able to feed this population. It is in this regard that precision agriculture matters most (Zhang et al. 2002).

Since Precision Agriculture (PA) aims to optimize and improve agricultural processes to ensure maximum product and to achieve this, there are many ways and we will focus on soil and its properties to the contribution of yield needed as mentioned above. Soil is said to be fundamental of all as long as you are talking about Agriculture Therefore, IoT is the best way of monitoring soil environment (here referred as soil properties) and alerting the users (agriculturalists) on real-time to take necessary actions. It is in this regard that this thesis will focus on monitoring soil properties using IoT and we know the quality and quantity of those selected properties towards predicting the most appropriate plant on that soil. To reach the expected goal, we need sensors, embedded system, database hosted on cloud for storage of data for future use, algorithms where machine learning is preferred (Ravesa et al,2021)

1.3 Problem Statement

Due to the increasing number of population and food is a basic need to every human being, the land to cultivate is not increasing instead it is decreasing due to shelter needs and land degradation because of climate change. Again the land cultivated for a long time loses its natural fertility and needs organic added manure to restore its fertility that is why we need to 1) To capture, store and monitor the soil properties/contents before applying that artificial manure 2) There is a need to first know what nutrients the land misses and type of manure and quantity needed hence to increase crop harvest and preserve soil without adding unnecessary fertile or which is inadequate to the land needs. 3) There is need of land use prediction where the model will be developed so that when you capture land soil the best appropriate crop is suggested. Also fertilizers not used deteriorates or hardens soil and can be washed away into the waters when it rains and this may affect living organisms in waters like fishes. After visiting website of RAB and meeting officials of RAB for the sake of more information on land use or if the related thesis was done, we found that no available data set that can be used the only thing they provided was reference standards of chemical parameters.

1.4 Study Objectives

The aim of this thesis is to provide real time data of a given land soil to farmers and decision makers for proper use and productivity of land. We need to collect and analyse data which will help farmers and Agriculturalists to grow crops with respect to land properties. Again there should be regulation of fertilizers either organic manure or inorganic by considering their long term effect.

1.4.1 General Objective

To design and implement IoT based monitoring system of soil properties.

1.4.2 Specific Objectives

1) To capture soil properties like PH, Temperature, humidity and NPK using sensors.

2) To store data from smart sensor in the database hosted on the cloud.

3) To make a prediction model which will predict croptype based to the soil properties currently available.

4) To provide a prototype for farmers who need to measure real time data for fertilization application.

The result of this thesis will be valuable to the farmers and Government authorities like RAB to take decisions on land preservation and crop rotation.

1.5 Hypotheses

Soil properties/nutrients monitoring using IoT and prediction using machine learning can help a farmer to know the right croptype in that land and apply fertilizers appropriately hence preserving soil and increasing productivity.

1.6 Study Scope

The study of this thesis focuses on capturing the specified soil parameters (PH, NPK, temperature, moisture) among many others, we chose these parameters because are very essential and common to any type of crop and sensors related to them are available on the market. The data captured are of chosen RAB garden at mulindi area because it is currently cultivated though we are in drought season. Sensors are calibrated based on test from standardized laboratory. We chose to focus on cultivated land instead of taking samples to any types of land which can work also.

1.7 Significance of the Study

This thesis focused on capturing soil properties, storing the captured data on database hosted on the cloud, monitoring of those properties and make a prediction model using machine learning to predict the best use of that land (here we specify the type of crop to plant on that land soil).The main contribution of this thesis to the community are many but we focused on few like:

Firstly, after consulting RAB officials, we found that there is no database so far where you can find the information of given land in Rwanda instead the use tradition method of taking sample in laboratory which has many drawbacks, so this system use IOT to capture information and keep them on database on cloud.

Secondly, the data captured and stored on cloud will be processed to predict the best use of land based on its current properties/nutrients. This means that it will benefit farmers by increasing productivity since they will know the soil nutrients available and how much fertilizer to add and know which crop is most adequate in data soil.

Finally, this system will mostly benefit agriculturalists and decision makers to decide land use based on vivid factors. It is very simple to test and know land properties it is just to take device on the field instead of using laboratory.

1.8 Organization of the Study

This thesis is organized in six chapters. Chapter one is the introductory chapter which involves the background of the thesis. Chapter two presents literature review explain the existing literature, what the gaps in this literature are and how you are going to solve this gaps.

Chapter three presents the research methodology explains the methods that are used in the implementation of this thesis. Chapter four presents system analysis and design discussing on the system circuit connections, algorithm and model. Chapter five presents results and analysis which discusses on the outcome and finally, chapter six draws conclusion and recommendations.

CHAPTER 2. LITERATURE REVIEW

In literature, we discuss thesis related works and gaps they highlighted while relating the contribution of this thesis.

Abhijit et al. (2019) proposed a smart system to support agricultural parameters using IoT. This system is explaining the best use of water in irrigation system based to the soil parameters like PH, Temperature, soil humidity and appropriate crops in a specific field. In the future work of this paper they suggested to use some more sensors like humidity sensor, carbon dioxide concentration sensors with more data analysis.

Collins et al. (2020) discussed on the mechanism to restore soil quality using biochar and brown coal waste. This article is about how to apply soil organic matter and brown coal waste to restore soil quality; they proposed future studies to focus on the long-term application these products and locally sourced alternatives.

Eunice et al. (2020) talked about agriculture conservation and integrated soil fertility management system for the maize production in the sub-humid and semi-arid regions of Kenya. This thesis was about soil and water management with integration of soil fertility management to enhance maize crop production, in their work they recommended to study the effects of both management practices on soil humidity during the long rain periods in the two regions.

Jashet al. (2019) suggested the system to optimize monitoring farming conditions using IoT. It is about better use of technology in farming to save resources (water and fertilizers) and increase yield, in their recommendation they emphases on IT infrastructures which must be in rural areas and its still challenge in under developing countries.

Mark et al. (2019) discussed the how soil biodiversity, fertility and function in a heathland restoration is affected by Long-term acidification of pH neutral grasslands.in their thesis they explained how field scale experimental with sulphur affected soil biodiversity, fertility, and function over a period of time (around 17 years). In this paper they recommended further work which is needed on application rates and timing as well as integration with grazing management.

Yanget al. (2014) Proposed to manage soil fertility in wheat-maize yield. This article is talking about the use of organic fertilizers and recycling of crop residues as very important and they compared different chemicals. They recommended to keep regulating the application of chemicals in the soil for environmental protection also.

Current research on IOT role in Agriculture are being conducted and it is shown that agriculture is a big field and needs more attention in order to maximize land productivity. Among many researchincluding the given ones in literature review, there is a general gap to bridge like database to host data for future analytics, creating model for prediction using machine because most of related work are talking about a situation for seeing individual situation like irrigation, monitoring one plant type like maize or tomatoes and applications are based on real-time data. Again I visited RAB during data collection and I met with soil department to help me and they gave me a contact person who provided only reference properties with some crops but no database, no automated system to collect data and store data. The main contribution of this work is that it has database hosted on the cloud where by a farmer can put a prototype (device) in garden with a certain crop and keep monitoring soil properties changes and this changes helps to know the rate of reduction of fertilizers and be able to add only needed fertilizer and if the farmer is not able to add fertilizer the system will suggest him the best appropriate crop can match with already available soil properties based to the prediction model. We requested to pattern with RAB and they promised to accept when we do it together based to seasons where we will need to take the whole season by monitoring a given sample to accredit it.

CHAPTER 3. RESEARCH METHODOLOGY

This is a set of systematic technique used in research. This simply means a guide to research and how it is conducted. It describes and analysis methods, throws more light on their limitations and resources, clarify their pre- suppositions and consequences, relating their potentialities to the twilight zone at the frontiers of knowledge (Professor Fidelis IfeanyiUgwuowo, 2016). They are many techniques and this thesis used data collection, data analysis to generate a model then we will be able to predict outcome and future use of information extracted from provided data.

Experimental research is primarily concerned with cause and effect. Here 7 the variables of interest are identified (i.e. the dependent and independent variables) and the researcher seeks to determine the effect of changes in the independent variables on the dependent variable.

In this thesis we used experimental research technique which is concerned with cause and effect and the variables of interest (dependent and independent variables) are identified. We used approach of prototyping whereby we collected primary data, make prediction model. We need to installed sensors in field/garden and collect soil properties then send data to the hosted database through wireless communication then analytics is done and model is used to predict the plant based on properties of that soil.

Data Collection: This is said to be the primary and important step for the research, in this thesis we collected primary data from the field by use of sensors then data captured are sent to cloud platform. This technique requires us to install sensor nodes in the field and sensor nodes should have capability to send captured data to cloud through internet connection. Data needed mentioned are NPK values of soil, soil PH, temperature and humidity/moisture.

Prototype: prototype is needed since we collect the primary data.it is made up of sensors, microcontroller and communication module.it is prototype not yet device because it will need modification before putting it on market. Here in the thesis it is used to test and show that work is validated and possible.

Data Analysis: This is taken as a tool used to extract useful information from raw data collected and help to make decisions. After collecting data we need to apply different algorithms of machine learning. The technique used helps us to generate a model. We choose predictive analysis because outcome of model help farmers to predict next use of land and kind of fertilizer to apply in the soil.

Data Interpretation/Reporting plan: This explains how to share information gathered to the end users like farmers and decision makers in the field of agriculture. To create user interface is planned so that farmer with smart phone can access data or written document and consultancy can be given to decision makers like RAB.

3.1 COMPONENT REQUIREMENTS

To make the system work properly the requirement of hardware and software tools are needed.

3.1.1 Hardware

3.1.1.1 PH sensor.

This sensor is inserted in 50 cm deep the soil to capture PH of soil.

3.1.1.2 Temperature sensor.

This sensor is inserted in 50 cm deep the soil to capture temperature of soil.

3.1.1.3 Humidity sensor.

This sensor is inserted in 50 cm deep the soil to capture humidity of soil.

3.1.1.4 NPK sensor.

This sensor captures three (3) parameters namely Nitrogen, Phosphorus and Potassium.

3.1.1.5 ARDUINO UNO

This microcontroller is the heart of the system, it is programmed using Arduino IDE. The data from sensors are sent to it and it processes them due to the codes and send data to the cloud based database for storage and analytics.

3.1.1.6 SIM 800L

This is GSM module which provides internet connection through cellular network.it has simcard slot. This gsm uses gprs capability and it is chosen because of its wide use country wide than any network infrastructure you can think, it is also cheap in terms of internet bundles and it is best fitting in agriculture field where my work belongs.

3.1.1.7 LCD.

This displays real time data on the field

3.1.1.8 Rechargeable battery.

This helps to supply the system in the field

3.1.2 Software

3.1.2.1 Arduino IDE

This software is used in Arduino uno by receiving live data from sensors and due to codes as programmed there is data cleaning to remove errors or nulls before sending to the cloud.

3.1.2.2 Database management system

This database is hosted on cloud to receive data from sensors by using gprs module for internet connectivity and the data stored is used for training model and testing and finally is used for further analysis in case they are needed.

3.1.2.3 Cloud application

Due to IOT era, we will use cloud for storage and future use of data for further analysis.

3.1.2.4 Machine Learning algorithm

This is the most part needed to analyse captured and collected data for predictions. In the next chapter we provide details of chosen and fitting algorithm due to the type of data and expected output.

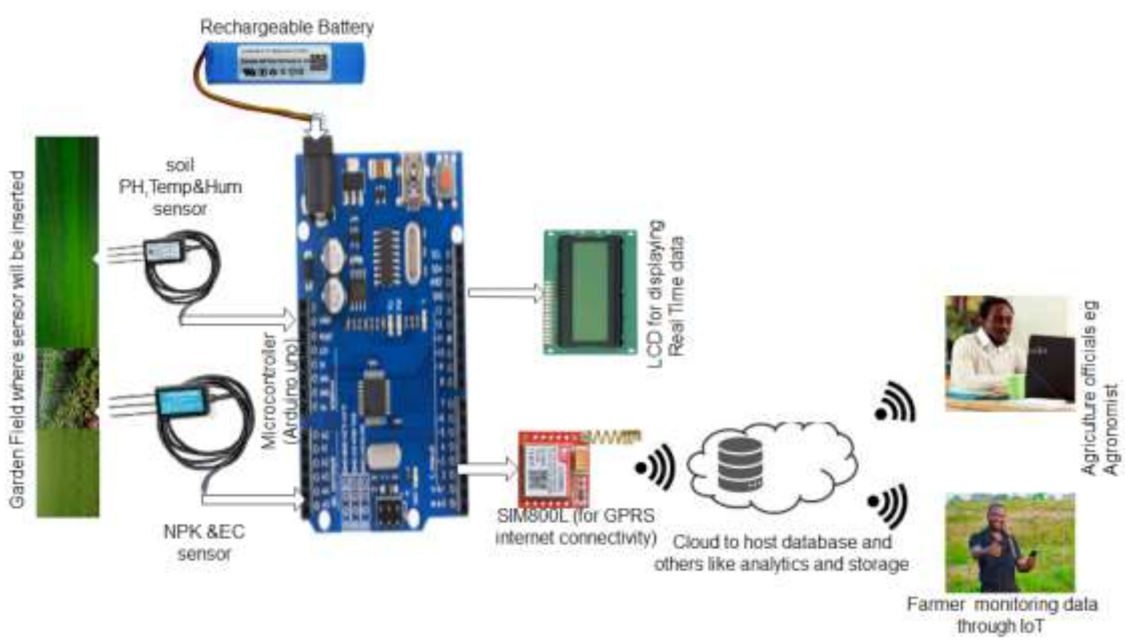

Figure 3.1 Proposed system architecture

In this system, sensors are in the field (garden to use) and they collect the specified data in explanations given that sensors are connected directly to Arduino and SIM 800L and the data are processed and sent to database which is hosted online. For Arduino and SIM 800L to send data to database, there is a need of internet connection. The data captured by sensor is displayed on LCD at the same time, the data is sent to the database and stored for prediction and for further use, For other explanations on how data from sensors are trained by model using classification regression that are used to generate the nearest crop type is detailed in the next chapter.

CHAPTER 4. SYSTEM ANALYSIS AND DESIGN

In this chapter we discussed system circuit connection, algorithm and model towards results.

4.1 System circuit diagram

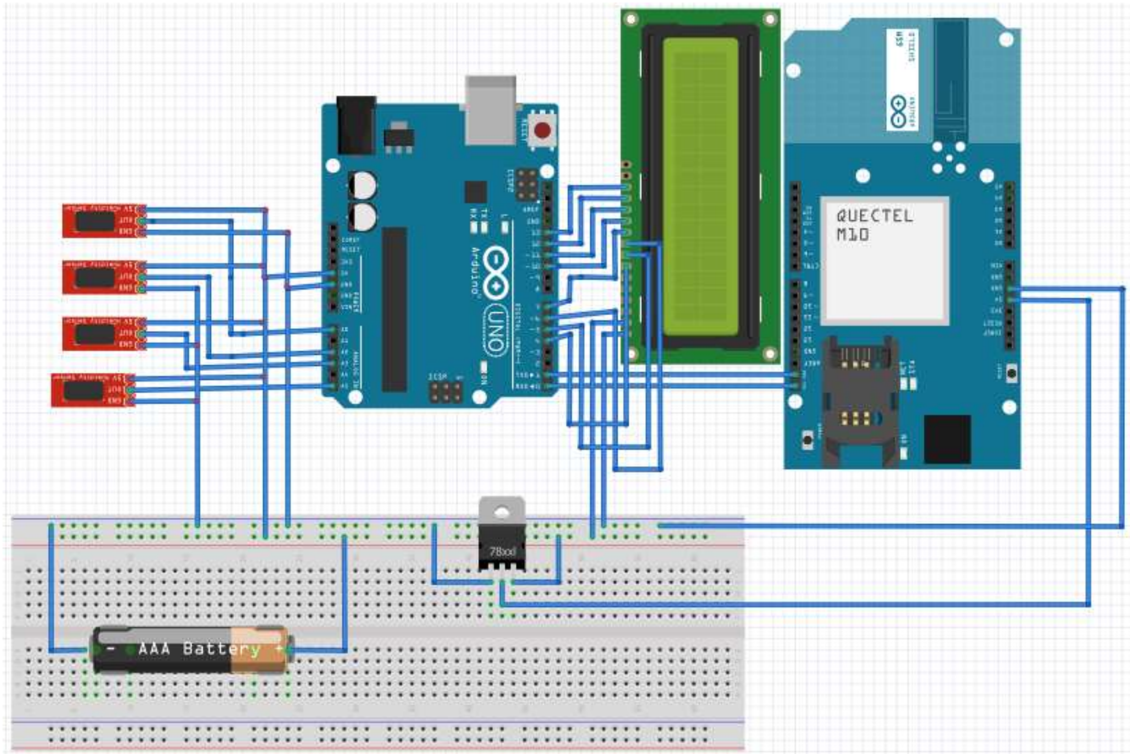

Figure 4.1 System circuit digram using fritzing software

This circuit is divided into 5 parts;part one is input part make up of four sensors to collect data which was mentioned earlier, part two is processing part made up of Atmega 328P which is in Arduino uno and its role is to receive data from sensors and process them according to the code, part three is communication module which is gprs to send data from microcontroller to the database through cellular system, part four is output of real-time data on lcd, part five is power supply system which has +5Vdc battery to supply sensors, microcontroller and lcd and 4.2v dc to supply gsm module.

4.2 System prototype

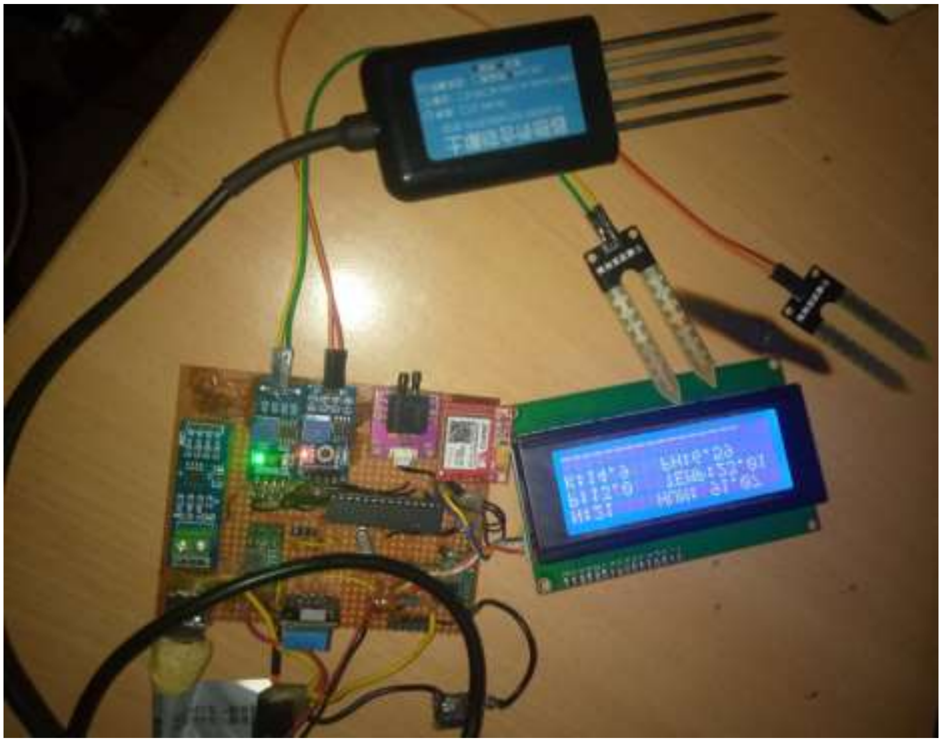

Figure 4.2 Prototype of the system

4.3 Algorithm to use

An algorithm is said to be a procedure or formula used to solve a problem. Therefore, we need codes to perform manipulations to achieve the results. More explanations and flow chart will be given in as we progress. Due to the type of output to predict which is categorical instead of numerical values, we found that Decision tree classifier model which is under classification algorithm is best fitting among others we trained like logistic regression and knn.

Classification is the type of supervised machine learning which is mostly fitting in a predictive modelling problem where class or labels are categorical instead of numerical values. I chose classification among others because it best fits in prediction where dependent variable (output) is a label (e.g. cereals).

Machine learning is a method of data analysis that automates analytical model building based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention (Vladimir Pavelčík and Erik Kuba, 2021)

Supervised learning is one of the most basic types of machine learning among others like unsupervised learning, in this work we chose supervised learning because we the model is trained using known labelled datasets while in unsupervised learning inputs are provided to the model (Professor R. Sathya and Professor Annamma Abraham, 2013).

Table 4.1 Sensor data as captured from the field by taking first 10 values on each crop type

ID	N	P	K	temperature	Humidity	ph	croptype
1	90	42	43	20.88	82	6.5	cereal
2	85	58	41	21.77	80.32	7.04	cereal
3	60	55	44	23	82.32	7.84	cereal
4	74	35	40	26.49	80.16	6.98	cereal
5	78	42	42	20.13	81.6	7.63	cereal
6	69	37	42	23.06	83.37	7.07	cereal
7	69	55	38	22.71	82.64	5.7	cereal
8	94	53	40	20.28	82.89	5.72	cereal
9	89	54	38	24.52	83.54	6.69	cereal
10	68	58	38	23.22	83.03	6.34	cereal
201	40	72	77	17.02	16.99	7.49	legume
202	23	72	84	19.02	17.13	6.92	legume
203	39	58	85	17.89	15.41	6	legume
204	22	72	85	18.87	15.66	6.39	legume
205	36	67	77	18.37	19.56	7.15	legume
206	32	73	81	20.45	15.4	5.99	legume
207	58	70	84	20.65	16.61	6.23	legume
208	59	70	84	17.33	18.75	7.55	legume
209	42	62	75	18.18	18.9	7.01	legume
210	28	74	81	18.01	18.31	8.75	legume
701	105	95	50	27.33	83.68	5.85	fruits
702	108	92	53	27.4	82.96	6.28	fruits
703	86	76	54	29.32	80.12	5.93	fruits
704	80	77	49	26.05	79.4	5.52	fruits

705	93	94	53	25.87	84.42	6.08	fruits
706	90	92	55	27.01	80.19	6.13	fruits
707	108	89	53	29.55	78.07	5.81	fruits
708	108	88	55	26.29	83.39	5.89	fruits
709	105	77	52	29.16	76.16	5.82	fruits
1200	61	68	50	35.21	91.5	6.79	vegetables
1201	58	46	45	42.39	90.79	6.58	vegetables
1202	45	47	55	38.42	91.14	6.75	vegetables
1203	39	65	53	35.33	92.12	6.56	vegetables
1204	31	68	45	42.92	90.08	6.94	vegetables
1205	70	68	45	33.84	92.85	6.99	vegetables
1206	68	62	50	33.2	92.76	6.98	vegetables
1207	34	65	47	23.49	93.71	6.83	vegetables

4.4 system Flowchart

This flowchart shows the flow of information from step one to the last in summary. Each step is detailed in next chapter

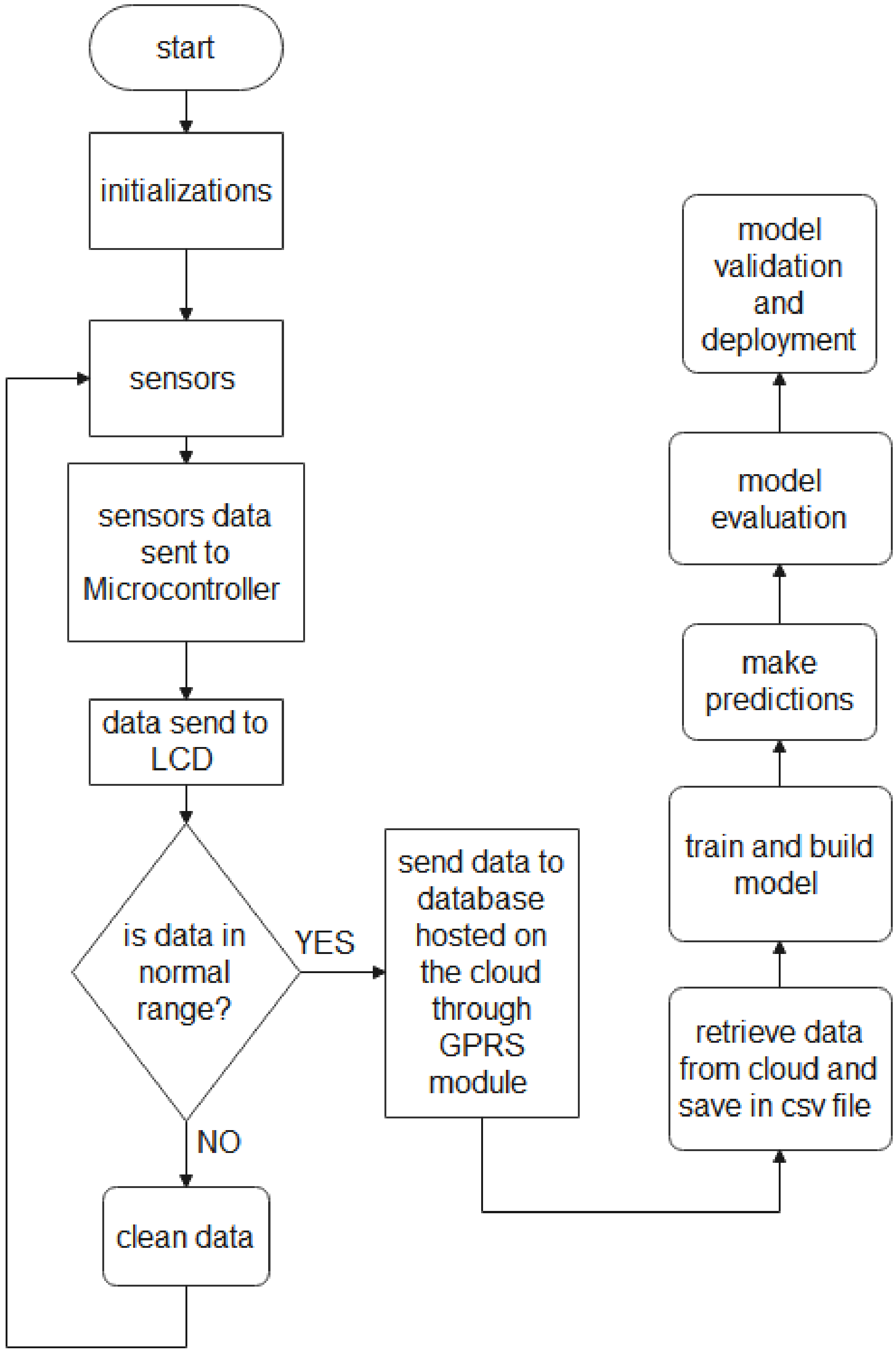

Figure 4.3 Flowchart of the proposed system algorithm

4.5 Dataset pre-processing and model training

There are different ways of python editor and chose to use jupyter note book which is user friendly and has many features.

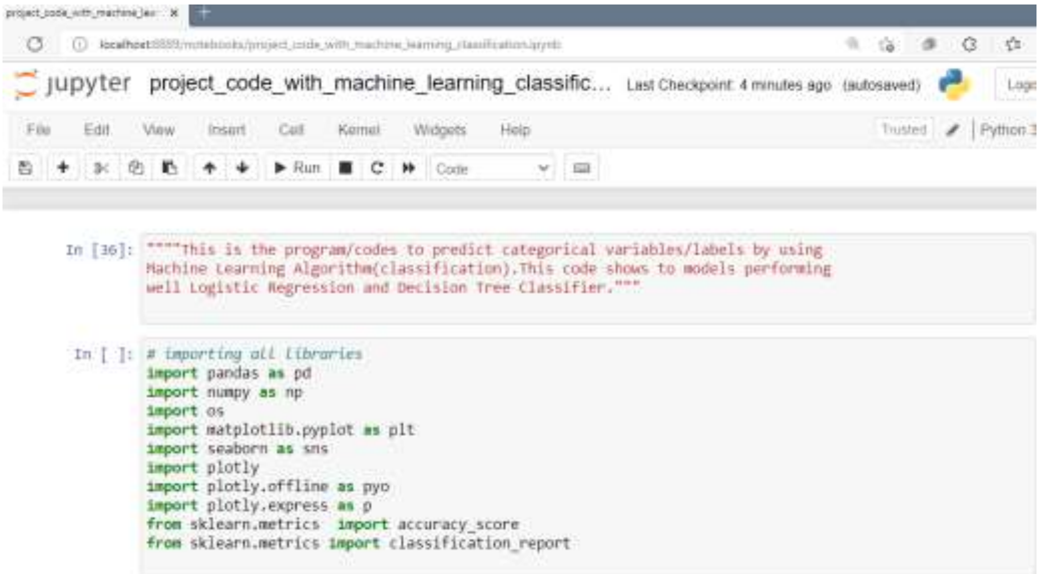

Figure 4.4 Main functions needed to visualize and manipulate our dataset.

```
#Load the data to use for training and testing,df means Dataframe
df=pd.read_csv('testingdata.csv')
df
```

	ID	N	P	K	temperature	humidity	ph	croptype
0	1	90	42	43	20.88	82.00	6.50	cereal
1	2	85	58	41	21.77	80.32	7.04	cereal
2	3	60	55	44	23.00	82.32	7.84	cereal
3	4	74	35	40	26.49	80.16	6.98	cereal
4	5	78	42	42	20.13	81.60	7.63	cereal
...	...	...	...	...	...	...	...	...
1284	1295	42	59	55	40.10	94.35	6.98	vegetables
1285	1296	43	64	47	38.59	91.58	6.83	vegetables
1286	1297	35	67	49	41.31	91.15	6.62	vegetables
1287	1298	56	59	55	37.04	91.79	6.55	vegetables
1288	1299	39	64	53	23.01	91.07	6.60	vegetables

1289 rows × 8 columns

Figure 4.5 Dataframe/dataset to train the model

```
# This scattering graph shows how features/independent variables are scatterd.here I used one but
#you can check on all features in any combination

sns.FacetGrid(plants, hue="croptype", height=10).map(plt.scatter, "temperature", "ph").add_legend()
```

```
<seaborn.axisgrid.FacetGrid at 0x1dd8a5021f0>
```

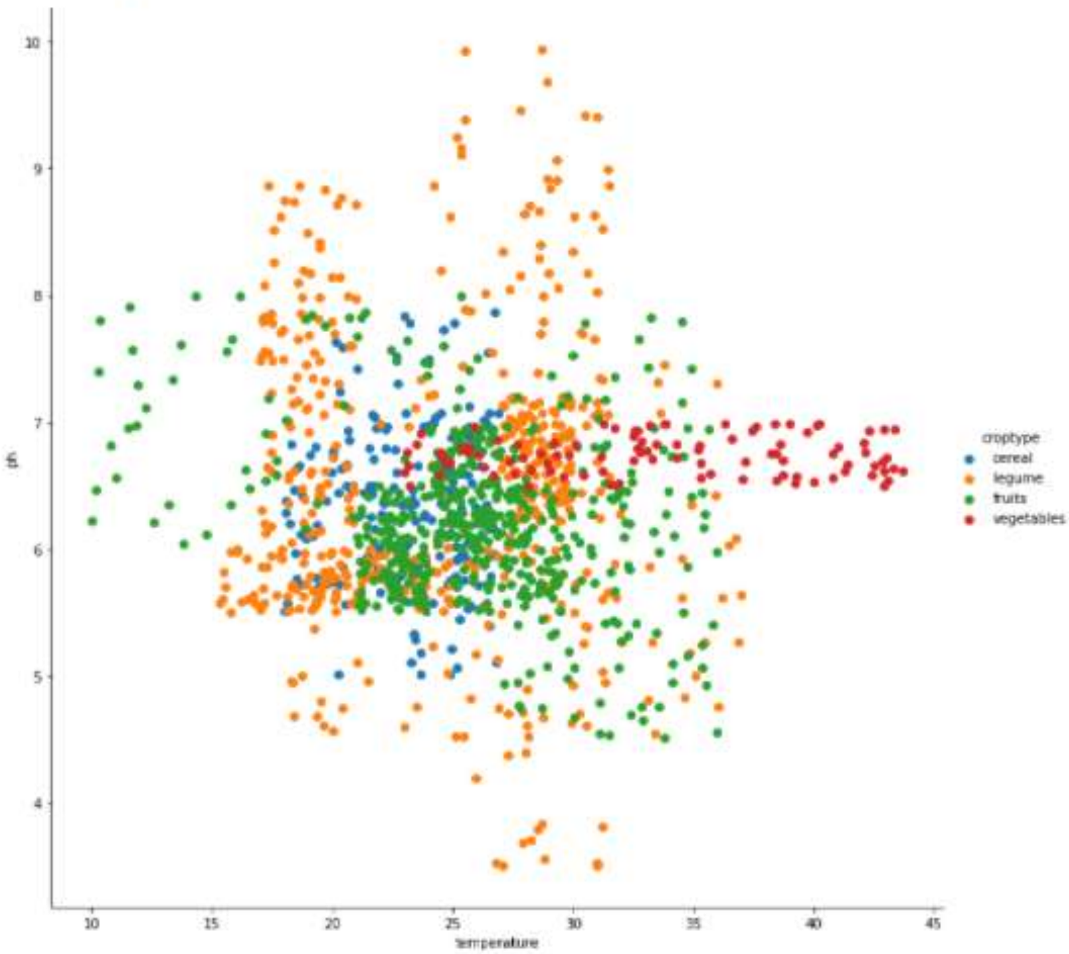

Figure 4.6Dataset scattering distribution based to dependent variables

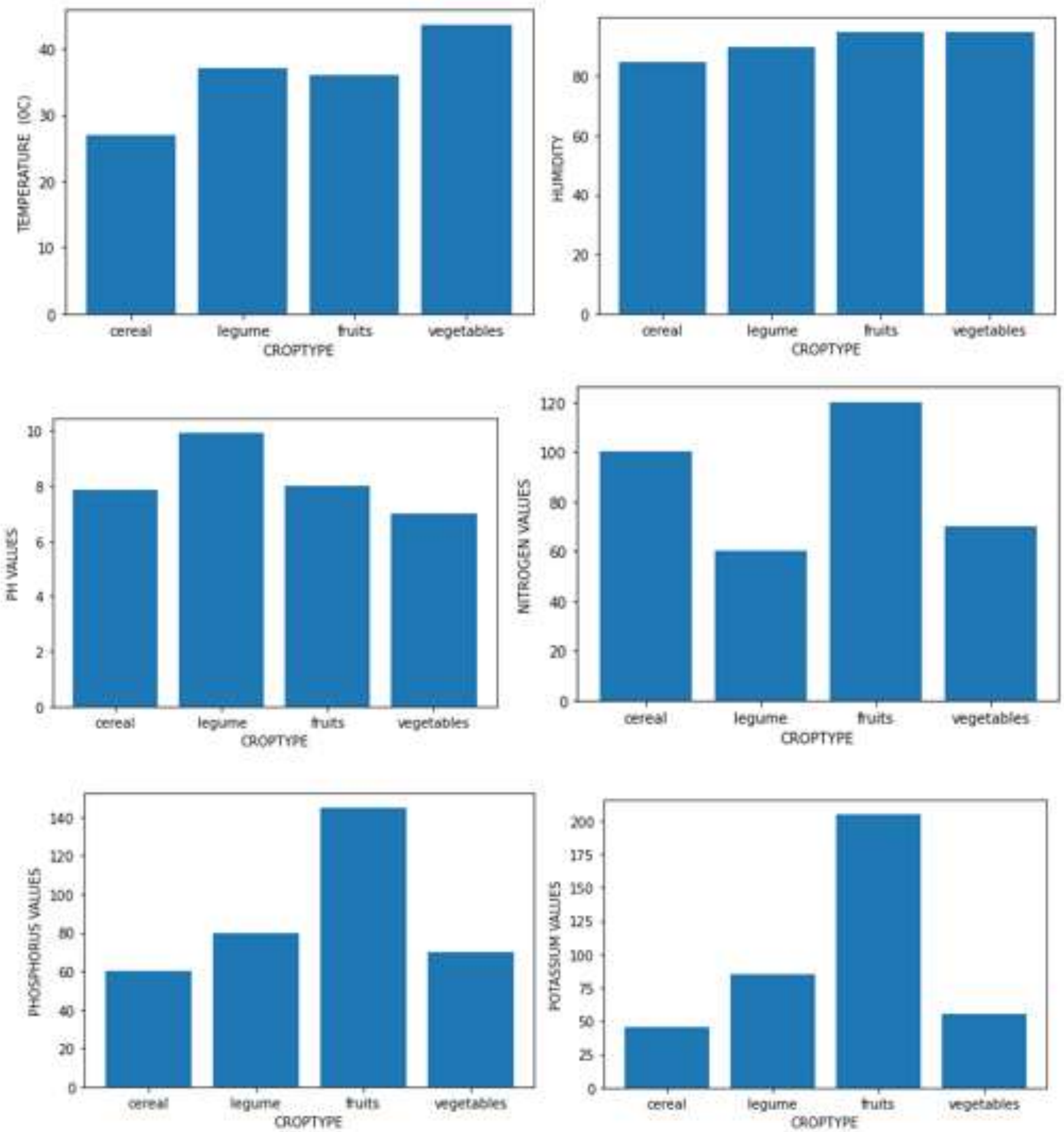

Figure 4.7Dataset visualization using bar graph

4.6 Model training

We have tried different models but here only two are shown.it means after trying many models we have to choose the one with highest accuracy (model validation)

4.6.1 Classification model using logistic regression algorithm

```
# model training

# train=70%   while   test=30%


from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x, y, test_size=0.30, random_state=42)
```

```
# logistic regression.
from sklearn.linear_model import LogisticRegression
LRmodel=LogisticRegression()
LRmodel.fit(x_train, y_train)
```

```
LogisticRegression()
```

```
prediction_LRmodel=LRmodel.predict(x_test)
```

```
# The accuracy of this model is 90.69% which is quit good
accuracy_LRmodel=accuracy_score(y_test,prediction_LRmodel)*100
```

```
print(accuracy_LRmodel)
```

```
90.69767441860465
```

4.6.2 Classification model using Decision tree classifier

```
#prediction by using second model which is Decision Tree
```

```
from sklearn.tree import  DecisionTreeClassifier
DTmodel = DecisionTreeClassifier()
DTmodel.fit(x_train, y_train)
```

```
DecisionTreeClassifier()
```

```
prediction_DTmodel=DTmodel.predict(x_test)
```

```
accuracy_DTmodel=accuracy_score(y_test,prediction_DTmodel)*100
```

```
# This  model od Decision Tree  is more accurate than Logistic Regression with 90.22%
print(accuracy_DTmodel)
```

```
99.2248062015504
```

4.7 Model validation

This is the process of choosing the model which is achieving the intended output among many models.it means after trying different models you can end up with no reliable model but for our case, we are choosing the Decision tree classifier which has accuracy of 99.2%.

```
from sklearn.tree import  DecisionTreeClassifier
DTmodel = DecisionTreeClassifier()
DTmodel.fit(x_train, y_train)
```

```
DecisionTreeClassifier()
```

```
prediction_DTmodel=DTmodel.predict(x_test)
```

```
accuracy_DTmodel=accuracy_score(y_test,prediction_DTmodel)*100
```

```
print(accuracy_DTmodel)
```

```
99.2248062015504
```

```
# Below shows  two output examples  when you path any random input to the model
```

```
x_DT=53,60,55,37,92,6.3
```

```
x_DT
```

```
(56, 60, 55, 37, 92, 6.7)
```

```
x_DT=np.array([[53, 60, 55, 37, 92, 6.3]])
x_DT_prediction=DTmodel.predict(x_DT)
```

```
x_DT_prediction
```

```
array(['vegetables'], dtype=object)
```

```
x_DT=91,42,43,20.88,82,6.7
```

```
x_DT
```

```
(91, 42, 43, 20.88, 82, 6.7)
```

```
x_DT=np.array([[91, 42, 43, 20.88, 82, 6.7]])
x_DT_prediction=DTmodel.predict(x_DT)
```

```
x_DT_prediction
```

```
array(['cereal'], dtype=object)
```

Figure 4.8 Prediction results of validated model

CHAPTER 5. RESULT AND ANALYSIS

In this chapter we discuss the outcome of the algorithm and its efficiency and if it can be trusted at what percentage to be realistic on the field.

5.1 system result

With validation of decision tree classifier which has accuracy of above 99%, we can say that the model is reliable and it can be deployed for use by end users.

```
print(classification_report(y_test,prediction_DTmodel))
```

```
              precision    recall  f1-score   support

      cereal       0.94      1.00      0.97        48
      fruits       1.00      0.99      1.00       143
      legume       1.00      0.99      0.99       167
  vegetables       1.00      1.00      1.00        29

    accuracy                           0.99       387
   macro avg       0.99      1.00      0.99       387
weighted avg       0.99      0.99      0.99       387
```

Table 5.1 Prediction summary each crop type and overall accuracy

This table above shows the precision, recall, score and support if each croptype and overall accuracy. For details we are taking one crop type which is cereal.

Precision indicates percentage of relevant items among retrieved items.it is taken as measure of quality. Precision is equal to true positives over all true positives including those not identified correctly. take an example for cereal, given the test size(support) of 48 which was named true positive ,45 out of 48 which is equal to 0.94 or 94% is well predicted while 13/48 was wrongly pressed named.

Recall indicates the percentage of relevant items that were retrieved.Recall is equal to the true positives over all samples that should be identified as positives. Take an example of cereal, given the support of 48, since it is 100% it means all items retrieved was relevant.is taken as measure of quantity.

F1-score is weighted average of precision and recall (94%+100%)/2=97% as it is in the table.

Support indicates the total samples. For example in the table above the support is mentioned 30% chosen as test data.

Accuracy indicates the number of correctly predicted observations to the total observations. The higher the accuracy the best the model. Since the accuracy of our model is 99%, it is best and it is validated to be used by beneficiaries.

5.2 Hypotheses and objectives approval

Based on the results, we can approve the hypothesis because prototype was made and it is capturing and displaying real time data and data is being sent to the database which means that a farmer can keep monitoring the reduction of soil nutrients/properties in the garden and be able to add fertilizers only when needed and estimate the quantity and the type of fertilizer missing, again the model was made using Decision tree classifier as machine learning algorithm to predict the best fit crop type on existing parameters.

All objectives was achieved as expected though few challenges was mentioned

5.3 Discussion

In this section we discuss the status of the use of IOT in agriculture in general and particularly in Rwanda. IOT was emerged in 1990s and its main practical functions was to expand the theory of Services Sciences, Management Engineering (SSME) and this lead it to infiltrate in all walks of life in different fields. Agriculture is one of the field interested in due to its importance for life.as mentioned in literature review, there are different ways of applying IOT in this field and the main purpose is to increase efficiency means small land, high crop productivity in short time(Kun et. al. 2012)

In agriculture most of applications as related to what is mentioned is about monitoring water, soil fertility of a given crop type and system reaction base on real time data for example when water decreases to a certain level irrigation system starts based to only temperature and moisture for a specific crop example maize or rice. Particularly in Rwanda there are very few application of IOT systems and most of them are based on irrigation and many of them including the mentioned ones in related work none with database to monitor not only real time data but also store on database for long term planning by government and none with predictions using machine learning.

The main issue is that the system (IOT) is new technology in Rwanda and IOT components to be used are very expensive, for example I bought only two sensors at $300 which means the beneficiaries cannot afford to pay easily. The other issue is about few knowledgeable people in the field but thanks to the government that placed ACEIOT to solve such problems and soon progress will occur.

This work will contribute more in academic field and in the Rwandan community on the level of farmers as well as on the level of decision makers like RAB since it combines collecting many (six) essential soil properties needed by a plant to grow and save those data on database hosted on the cloud and provide a model for predictions of best use of land using machine learning.

CHAPTER 6 : CONCLUSION AND RECOMMENDATION

In this work we have presented the results of IoT based monitoring system of soil properties and crop type prediction using machine learning algorithm. The prototype allows real time measurement of data and storing those data on cloud database which was used in training and testing model.

The most challenge faced was lack of sensors on time because we did not find them on local market of Rwanda and we needed to order from outside. Again the issue of calibration of those sensors which are new to the market is very difficult.

The other challenge is that there is a need of advanced IT skills which was difficult to handle tasks related to database system which like creating user interface so that the end user can monitor data in the database on back end where he can only monitor and visualize information.

The challenge of deploying model which is created using python was a problem because most of deploying is using apache database management system and it is not easy to accomplish this task of model deployment.

We recommend that modules related to IT should be added in wisenet as well as in ECS option and many practical skills are needed.

We recommend the center to provide financial support to conduct research very early so that a student has room of doing deep research which can benefit community rather than academic field, we have noticed that there are some projects (mostly those related to agriculture or with a need of big data collection) that need more time than what we really got.

REFERENCES

AbhijitPathaka, Mohammad AmazUddina, Md. JainalAbedina, Karl Anderssonb, RashedMustafac, Mohammad Shahadat Hossainc(2019). *IoT based Smart System to Support Agricultural Parameters*.Halifax, Canada.

Collins Amoah-Antwi,StevenF(2020).*Restoration of soil quality using biochar and brown coal waste*.Science of total environment.

Eunice A. Mutuku, Dries Roobroeck, Bernard Vanlauwe, Pascal Boeckx, Wim M. Cornelis (2020) *Maize production under combined Conservation Agriculture and Integrated Soil Fertility Management in the sub-humid and semi-arid regions of Kenya*. Field Crops Research.

https://borgenthesis.org/sustainable-agriculture-in-rwanda/

https://fortuneofafrica.com/rwanda/challenges-faced-by-the-agriculture-sector/

JashDoshi, Tirthkumar Patel, Santosh kumar Bharti (2019).*Smart Farming using IoT, a solution for optimally monitoring farming conditions*. Coimbra, Portugal.

JaydeepKarandikar,Tony Schmitz and Scott Smith (2020).*Logistic classification for tool life modelling in machining*. University of Tennessee, Knoxville, USA.

Keyur K Patel and Sunil M Patel (may2016).*Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges*. Vadodara, Gujarat, India.

Kun Han,Shurong Liu,Dacheng Zhang and Ying Han (2012). *Initially Researches for the Development of SSME under the Background of IOT*.Hebei Province, China.

Mark Tibbett, Marta Gil-Martínez, Tandra Fraser, Iain D. Green, Sarah Duddigan,

Priyanka Rajendra and ShahramLatifi (2021*). Prediction of diabetes using logistic regression and ensemble techniques*. UNLV, Las Vegas, Nevada, USA.

Professor Fidelis IfeanyiUgwuowo (2016).*Fundamentals of research methodology and data collection*.University of Nigeria, Nsukka.

Professor R. Sathya and Professor Annamma Abraham (2013*). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification*. International Journal of Advanced Research in Artificial Intelligence. Bangalore, India.

Vinicius H. De Oliveira, Karsten Raulund-Rasmussen, Tom Sizmur, Anita Diaz.(2019*). Long-term acidification of pH neutral grasslands affects soil biodiversity, fertility and function in a healthland restoration*. Catena.

Vladimir Pavelčík and Erik Kuba (2021) *Application of basic machine learning algorithms in railway brake disc temperature prediction*. 4th International scientific conference on sustainable, modern and safe transport.

YANG Xue-yun, SUN Ben-hua and ZHANG Shu-lan(2014).*Trends of Yield and Soil Fertility in a long-term Wheat-maize System*. Journal of Integrative Agriculture.

APPENDIX

Table Appendix 1. Interpretation values of chemical parameters according to Landon (1991)

Analysis and unity	Average values of the analysis	Soil class and / or qualification
pH water	>8,5	Very high (alkaline soil)
	7,0 - 8,5	High (weakly alkaline)
	5,5-7,0	Medium (weakly acidic)
	<5,5	Weak and very weak (acidic)
CEC (meq/100g)	>40	Very high
	25-40	High
	15-25	Average
	5-15	Low
	<5	Very low
% base saturation (ratio in% of exchangeable bases and CEC)	>60	High
	20-60	Average
	<20	Low
Calcium	>10	High

	<4	Low/weak
Magnesium	>4	High
	<0,5	Low
Potassium	>0,6	High
	<0,2	Low
Sodium	>1	High
	<1	Low
Organic carbon in%	>10	High
	4-10	Medium
	<4	Low
Total nitrogen (Kjeldahl) in%	>0,5	High
	0,2-0,5	Medium
	<0,2	Low
Available phosphorus in ppm (depending on the method) if it is the Bray method (recommended for acidic soils)	>50	High
	50-15	Medium
	<15	Low

Table Appendix 2 Standard values on selected parameters which will be taken as reference or sensor calibration

ID	Plant	PH	N	P	K	Temperature°C	Humidity%
1	Maize	6.0-6.5	14	23	14	21 -28	40
2	soybean	5.8-6.2	1	4	2	26.5-30	30
3	Beans	6.0-6.8	10	5	16	10-21	29
4	Flowering plant		6	2	4	13-20	28
5	Wheat	4.4-6.9	4.9	1.0	8.9	25-30	19
6	Carrot	5.5-7.0	0/5	10/15	10/15	13-21	20
7	Onion	5.8-7.0	6	4	4	13-24	21
8	cabbage	6.0-7.5	10	10	10	13-24	22
9	Rice	6.2-6.7	20	10	10	30-32	23
10	cassava	6-7	12	11	18	25 - 29	24
11	sorghum	6-7.5	4	2	0	25–28	25
12	banana	4.4-5.2	8	10	8	20- 27	25
13	groundnuts	5.7-6.3	17	34	54	20- 30	25
14	sunflower	6.0-7.5	10	5	5	21-26	25
15	sugarcane	5.0-8.5	15	8	8	21- 27	25
16	Yams	5.5-6.5	20	10	10	25 – 30	21
17	Tea	4.2-5.0	25	5	5	21- 29	21