

**EVALUATION OF AN INTELLIGENT AGRO-CLIMATE
DECISION SUPPORT TOOL FOR SMALL-SCALE
FARMERS:**

**An integration of mobile phones, smart sensors, and
indigenous knowledge**

By

NALEDI PORTIA THOTHELA

Submitted in fulfilment of the requirements for the degree:

MASTER OF INFORMATION TECHNOLOGY

In the Department of Information Technology

Faculty of Engineering, Built Environment and Information Technology

Central University of Technology, Free State

Supervisor: **Prof. ED Markus**

Co-supervisor: **Prof. EM Masinde**

Co-supervisor: **Prof. AMI Abu-Mahfouz**

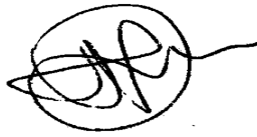
Bloemfontein

2021

Declaration

I, Naledi Portia Thothela, hereby declare that this research project which has been submitted to the Central University of Technology, Free State for the degree MASTER OF INFORMATION TECHNOLOGY, is my own independent work, complies with the Code of Academic Integrity and conduct, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology, Free State, and has not been submitted before by any person in fulfilment (or partial fulfilment) of the requirements for the attainment of any qualification.

Student Signature:



Date: 14 July 2022

Acknowledgements

First and foremost, I would like to acknowledge the one who has made all this possible for me, my Heavenly Father, who has seen me through all the years of my studies and provided for my every need. I would not be here if it was not for His grace and mercy. He has made everything work for my good and has blessed me with all these people that I am about to acknowledge below.

My most sincere appreciation to my supervisors Prof. Elisha Markus, Prof. Adnan Abu-Mahfouz, and Prof. Muthoni Masinde for their support, their continuous guidance, and all the invaluable insight and intense knowledge imparted. I have been highly honoured and privileged to have had the opportunity to work with supervisors of their calibre. Special thank you to Prof Muthoni Masinde for believing in me and for reminding me who I am.

Thank you to the Department of Information Technology for affording me the opportunity to be a part of this course. My undying gratitude also goes to Central University of Technology and CSIR for the financial support.

My appreciation goes to Dr Tafadzwanashe Mabhaudhi, from The University of KwaZulu-Natal in Pietermaritzburg, for his commitment in assisting me with my project, that includes carrying travel expenses to Pietermaritzburg, availing all resources necessary and opening doors for me to be able to work with the farmers at Swayimane.

Thank you to Masello Ramathe and Azikiwe Nobanda for giving me their unfailing love and support through the toughest time in my life. I could not ask for better friends. Finally, thank you to Princess Thothela, my daughter, for being my cheerleader and my consistent and constant support for all my study years. I know it has not been easy for her through this time, but I am grateful that she encouraged me through the toughest of times and unfailingly reminded me that it is all worth it.

Abstract

Food security in Africa and the rest of the globe has come under tremendous threat. This means that agriculture, being the main driving force behind many economies is under threat. Seventy percent of the food produced in Sub-Saharan Africa comes from the rain-fed small-scale agriculture. This agricultural sector happens to be the most devastated by any disasters experienced in the agricultural sector such as floods, drought, and other extreme climatic conditions. Evidence show that Africa has accounted for the most global droughts. Droughts have also become increasingly common in recent years. For example, out of 106 global droughts observed between 2010 and 2019, eight occurred in Africa in 2019. Catastrophically, these droughts affected 66 countries and impacted 690.2 million people, 9.3 million of whom were in Africa. The dire case for Africa can be attributed to lack of timeous and relevant early warning systems which is a result of limited resources in the small-scale agricultural societies. Indigenous Knowledge is currently still the trusted prediction tool being used by small-scale farmers for their day-to-day operations and strategic agricultural decision support. Climate change and global warming have however rendered this knowledge unreliable and unpredictable.

There is a plethora of scientific models, decision support tools and predictive methods that have been researched, all in the quest to find the solutions to the declining food production in the agricultural sector. Furthermore, much investment has been put behind research in the agricultural sector all over the world. The relationship between the Indigenous Knowledge (IK) and scientific methods has been substantially debated, but researchers have concluded that science and indigenous knowledge complement each other rather than compete. Limitations to both IK and modern scientific tools has led to this study which aimed to investigate an integration of Indigenous knowledge, mobile phone, and smart sensor technology with intelligence, in one system that can play a role in assistive decision support tool. In order to enhance and extend IK, the concept of installing smart sensors in the field with the view to measure humidity, Phosphorous, Potassium and Nitrogen concentration, and for detection of possible infestation of crops.

For the theoretical framework, this study adopted and adapted the ITIKI framework as a foundation. ITIKI is a drought early warning system developed for small-scale farmers

in Sub-Saharan Africa. The system uses input from observed indigenous knowledge indicators according to various farming activities and uses the fuzzy inference system for an assumption output. Weather forecast and historic climate data forms part of the system input. The necessary and relevant messages are conveyed to the small-scale farmers through mobile phone. The forecast provided by the ITIKI framework gives valuable information in making decisions on whether to plant, when to plant and even how and what to plant. To enhance the intelligence of ITIKI, this study investigated the application of machine learning algorithms to assist small-scale farmers in a crop selection process at the planting stage. The study explored machine learning classification models to find the best possible model which resulted in an Agro-climate Decision Support tool.

This tool was developed to be able to assist farmers in making a crop selection decision prior to planting. This tool was developed using the blueprint for integrating the IK and scientific model. Machine learning (ML) was used to determine the model to be integrated with the IK for developing the intelligence of the system. Climatic data, together with edaphic data ran through different ML algorithms to determine the best algorithm. The best model was used to select the best crop to plant based on the edaphic and climatic data. The models were tested and evaluated using the Jupyter notebook's different metrics. The system compares the results of the ML with the observed IK to determine the crop to be planted. In a case where the results prove contrary, a decision requires a crop scientist intervention. The evaluations proved that the latter was possible because the climatic and edaphic conditions could affect the accuracy of the ML model, and the confusion matrix results suggested different crops at times. In conclusion, the study developed a cropping decision support tool for integration of Indigenous knowledge, sensor technology data, machine learning and mobile phone technology in a form of the Intelligent Agro-climate decision support system to be used by small-scale farmers.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks
API	Application Programming Interface
DSS	Decision Support Systems
FIS	Fuzzy Inference System
FN	False Negative
FP	False Positive
GIS	Geographic Information Systems
GPS	Global positioning system
ICT	Information and communication Technology
IK	Indigenous knowledge
IoT	Internet of Things
ITIKI	Information Technology and Indigenous Knowledge with Intelligence
K	Potassium
KNN	K-Nearest Neighbour
ML	Machine learning
MMS	Multimedia Messaging Systems
N	Nitrogen
P	Phosphorous
RBF	Radial Bias Function
RF	Random Forest
SMS	Short Messaging Systems
SVC	Support Vector Classifier
TN	True Negative
TP	True Positive

Table of Contents

Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
Abbreviations.....	v
Table of Contents.....	vi
List of Tables.....	x
List of Figures.....	xi
Chapter 1: Introduction.....	1
1.1 Background.....	1
1.2 Problem Statement.....	3
1.3 Objectives of the study.....	4
1.4 Research Methodology.....	5
1.5 Contribution to the knowledge.....	6
1.6 Research publications from the study.....	6
1.7 Limitations of the study.....	7
1.8 Outline of the dissertation.....	7
Chapter 2: Literature Review.....	8
2.1 Introduction.....	8
2.2 Overview of Mobile phone technology.....	8
2.2.1 Mobile technology in Sub-Saharan Africa.....	9
2.2.2 Mobile technology in Agriculture.....	11

2.3	Integration of Indigenous Knowledge and ICTs in Small-scale Farming in Africa	12
2.3.1	Indigenous knowledge.....	12
2.3.2	ICTs in Small-scale Agriculture.....	12
2.3.3	Integration of indigenous knowledge.....	14
2.4	Decision Support Systems for Small-scale Farmers.....	16
2.5	Application of Artificial Intelligence for Cropping Decision Support.....	18
2.5.1	Application of Machine learning for cropping decisions.....	18
2.5.2	Fuzzy Inference System.....	19
2.5.3	Application of Fuzzy Inference System to cropping decisions.....	20
2.6	Discussion and comparison of considered technologies.....	21
2.7	Conclusion.....	26
Chapter 3: Research Methodology.....		28
3.1	Introduction.....	28
3.2	Research design.....	28
3.2.1	Machine learning models.....	28
3.2.2	Framework Development.....	29
3.2.3	Software prototyping.....	30
3.2.4	Case Study.....	31
3.3	Data collection.....	33
3.3.1	Indigenous Knowledge Data.....	33

3.3.2	Edaphic Data	37
3.3.3	Weather Data	38
3.4	Data generation	39
3.5	Data analysis	40
3.5.1	Demographics of the respondents	40
3.5.2	Testing the questionnaire reliability	43
3.6	Evaluation procedures	47
3.7	Conclusion.....	47
Chapter 4: Modelling of the Intelligent Agro-climate Decision Support system.....		48
4.1	Introduction	48
4.2	Machine learning.....	48
4.2.1	Data collection.....	49
4.2.2	Data Analysis.....	51
4.2.3	Data pre-processing	61
4.2.4	Choosing an algorithm	62
4.2.5	Training the dataset	63
4.3	Findings.....	79
4.4	Conclusion.....	79
Chapter 5: Simulation results and discussion		81
5.1	Introduction	81
5.2	Simulations Results	81

5.2.1 Accuracy score	81
5.2.2 Confusion Matrix.....	83
5.2.3 Classification report.....	86
5.3 Process of elimination	88
5.4 Conclusion.....	89
Chapter 6: Conclusion and Future work	90
6.1 Conclusion.....	90
6.1.1 Summary of the study.....	90
6.1.2 Summary of the functionality of the IACDSS	92
6.2 Future work	94
References.....	95
Appendix A	103
Sample code	103
Decision Tree algorithm	103
K-nearest neighbours algorithm.....	103
Data normalization process.....	104
Random Forest algorithm	104
Gradient boosting algorithm.....	104
Support vector machine algorithm.....	105
Appendix B	106
Sample questionnaire.....	106

List of Tables

TABLE 2.1 COMPARISON OF EXISTING TECHNOLOGIES USED.....	24
TABLE 3.1 KZN INDIGENOUS KNOWLEDGE WEATHER INDICATORS	35
TABLE 3.2 EDAPHIC AND CLIMATIC CONDITIONS AFFECTING CROP SELECTION.....	37
TABLE 3.3 CLIMATIC DATA COLLECTED FOR PIETERMARITZBURG.....	38
TABLE 4.1 PLANTING CHART FOR KWAZULU-NATAL.....	50
TABLE 4.2 CONTENTS OF THE CROP SIMULATION FILE.....	51
TABLE 4.3 STATISTICAL DESCRIPTION OF THE CROP SIMULATION FILE	52
TABLE 4.4 MACHINE LEARNING LIBRARIES USED	64
TABLE 5.1 SUMMARY OF ACCURACY SCORES	83
TABLE 5.2 SUMMARY OF CLASSIFICATION REPORTS.....	87

List of Figures

FIGURE 2.1 EVOLUTION OF MOBILE PHONE TECHNOLOGY	9
FIGURE 2.2 STAGES IN FUZZY MODELLING	20
FIGURE 2.3 INTELLIGENT MOBILE APPLICATION FOR CROP DISEASE DIAGNOSIS	21
FIGURE 2.4 ITIKI: INTEGRATED DROUGHT EARLY WARNING SYSTEM FRAMEWORK	22
FIGURE 3. 1 MACHINE LEARNING PROCESS	29
FIGURE 3. 2 THE INTERFACE OF THE MOBILE APPLICATION PROTOTYPE DEVELOPED	31
FIGURE 3. 3 MAP OF THE MUNICIPALITIES IN KZN MIDLANDS	32
FIGURE 3. 4 MAP OF UMSHWATHI MUNICIPALITY WARD 8.....	32
FIGURE 3. 5 CLOSER LOOK AT UMSHWATHI WARD 8 (LOCATION OF SWAYIMANE)	33
FIGURE 3. 6 BAR CHART OF THE GENDER DEMOGRAPHICS.....	40
FIGURE 3. 7 AGE GROUPS OF THE FARMERS	41
FIGURE 3. 8 LEVEL OF EDUCATION	42
FIGURE 3. 9 LEVEL OF EXPERIENCE IN FARMING	43
FIGURE 3. 10 RESPONSES FOR QUESTION 1	44
FIGURE 3. 11 RESPONSES FOR QUESTION 2.....	44
FIGURE 3. 12 RESPONSES FOR QUESTION 3	45
FIGURE 3. 13 RESPONSES FOR QUESTION 4.....	45
FIGURE 3. 14 RESPONSES FOR QUESTION 5	45
FIGURE 3. 15 DESCRIPTIVE STATISTICS FOR THE DATASET	46
FIGURE 3. 16 RELIABILITY RESULTS FOR THE DATASET	46
FIGURE 4.1 HEAT MAP OF THE CONTENTS OF THE CROP SIMULATION FILE	53
FIGURE 4.2 COUNT PLOT OF THE DATA DISTRIBUTION	54
FIGURE 4.3 COMPARISON BETWEEN TEMPERATURE AND PH.....	54

FIGURE 4.4 COMPARISON BETWEEN TEMPERATURE AND HUMIDITY	55
FIGURE 4.5 HEATMAP OF CORRELATION BETWEEN FEATURES.....	56
FIGURE 4.6 HISTOGRAM OF EACH OF THE FEATURES.....	57
FIGURE 4.7 DENSITY PLOT OF EACH FEATURE	58
FIGURE 4.8 JOINT PLOT OF THE RAINFALL AGAINST HUMIDITY SHOWN FOR 3 CROPS	59
FIGURE 4.9 BOX AND WHISKERS PLOT FOR EACH FEATURE	60
FIGURE 4.10 MACHINE LEARNING PROCESS FOR ALGORITHM SELECTION.....	64
FIGURE 4.11 FEATURE IMPORTANCE ACCORDING TO DECISION TREE CLASSIFICATION ...	65
FIGURE 4.12 DECISION TREE ACCORDING TO THE DECISION TREE CLASSIFIER.....	67
FIGURE 4.13 CLASSIFICATION REPORT FOR THE DECISION TREE CLASSIFIER	68
FIGURE 4.14 CONFUSION MATRIX FOR KNN CLASSIFICATION	70
FIGURE 4.15 CLASSIFICATION REPORT FOR KNN CLASSIFICATION.....	70
FIGURE 4.16 CLASSIFICATION REPORT FOR KNN	71
FIGURE 4.17 CONFUSION MATRIX FOR RANDOM FOREST CLASSIFIER	72
FIGURE 4.18 CLASSIFICATION REPORT FOR RANDOM FOREST CLASSIFIER.....	73
FIGURE 4.19 CONFUSION MATRIX FOR THE GRADIENT BOOSTING CLASSIFIER.....	74
FIGURE 4.20 CLASSIFICATION REPORT FOR GRADIENT BOOSTING CLASSIFIER	74
FIGURE 4.21 CLASSIFICATION REPORT FOR SVM LINEAR KERNEL	75
FIGURE 4.22 CONFUSION MATRIX FOR THE SVM LINEAR KERNEL.....	76
FIGURE 4.23 CONFUSION MATRIX FOR SVM RADIAL BIAS FUNCTION CLASSIFIER.....	77
FIGURE 4.24 CLASSIFICATION REPORT FOR SVM RADIAL BIAS CLASSIFIER	77
FIGURE 4.25 CONFUSION MATRIX FOR SVM POLYNOMIAL CLASSIFIER.....	78
FIGURE 4.26 CLASSIFICATION REPORT FOR SVM POLYNOMIAL CLASSIFIER	78
FIGURE 5.1 CONFUSION MATRIX FOR DECISION TREE CLASSIFIER.....	83
FIGURE 5.2 CONFUSION MATRIX FOR GRADIENT BOOSTING CLASSIFIER.....	84

FIGURE 5.3 CONFUSION MATRIX FOR KNN CLASSIFIER.....	84
FIGURE 5.4 CONFUSION MATRIX FOR RANDOM FOREST.....	85
FIGURE 5.5 CONFUSION MATRIX FOR SVM LINEAR CLASSIFIER.....	86
FIGURE 6. 1 THE INTELLIGENT AGRO-CLIMATE DECISION SUPPORT SYSTEM	93

Chapter 1: Introduction

1.1 Background

Information and Communication Technologies (ICT) in agriculture, may generally include technologies such as Global Positioning System (GPS), robotic systems, sensor technology and Geographic Information Systems (GIS). There is a plethora of these kind of technologies already developed and put to test in an effort to support agricultural practices and operations in order to improve crop production [1]. It should not come as a surprise therefore that significant research effort and resources have been put behind agricultural innovation. Agriculture is a mammoth economic driver in most countries and has the highest impact on the world's food security. Of all the food produced in Sub-Saharan Africa, 80% is produced in the informal rain-fed small-scale farms. According to the World Disaster report of 2021 [2], out of 106 of the world's droughts reported between 2010 and 2019 affecting 66 countries and impacting 690.2 million people, 8 occurred in 2019 only in Africa, and in 2019, droughts in Africa impacted 9.3 million people. The global Drought Observatory Analytical report observed that in the year 2021, the standard precipitation index analysis for Sub-Saharan Africa was lower, indicating that dryer conditions were experienced [3]. It is evident therefore, that Africa is leading in terms of global droughts in comparison to the rest of the world [4]. What is more striking about these statistics is that these droughts have been reported to have been the cause of about 99% of reported deaths that occurred in Africa. Evidently, Africa is the worst affected by droughts, and small-scale farmers are the most affected by the most adverse disasters like these, due to lack of timeous warnings and relevant information [5].

It is important to differentiate small-scale farmers from the rest of the agricultural sector. "A small-scale farmer is one whose scale of operation is too small to attract the provision of services he/she needs to be able to significantly increase his/her productivity", according to Kirsten *et al.*, [6]. These farmers do not have access to equipment and machinery that can make their work easier. In fact, their work is labour intensive, and they rely on their human strength. They also have no access to any kind of ICT assistive tools due to the high costs related to such. Access to information such as weather forecast is another limitation for the small-scale farmer. Small-scale farmers mostly rely on their

indigenous knowledge as a form of forecast and the basis for their daily decision support operation [7].

Indigenous knowledge (IK) refers to the native knowledge or local knowledge that is normally passed on from generation to generation. This is normally based on the extrinsic environmental behaviours, including seasonal changes, meteorological behaviours, astrological behaviours, flora and fauna behaviours, insect behaviours and spiritual beliefs. This knowledge is often not documented and is informal. There has been debates around IK in comparison with scientific methods, but research has shown that indigenous knowledge, though different from scientific methods, may very well be complementary to scientific methods, and that collaboration between the two can provide an enormous contribution to technological innovation[8]–[10]. Indigenous knowledge has already formed part of some relevant and appropriate innovative technologies already created [11]. Including IK as part of technologies developed, provides an opportunity for collection, preservation, and documentation of the indigenous knowledge for future references.

The integration of indigenous knowledge into mobile technology may seem unconventional yet could provide the ideal solution for the following reasons. Mobile phone technology is playing a colossal role in bridging the digital divide in Africa [12], i.e. the gap that exists between the demographic regions that have no access to modern ICT and internet and those that do not. The mobile technology in comparison to other ICTs is more affordable, user-friendly, and less complicated in a sense that it has less infrastructure requirements. The mobile technology footprint in Africa stood at higher than 63% in 2014[8]. In Africa, mobile phone technology overtook the use of landlines [12] to become the preferred mode of communication. In comparison to other available technologies in terms of performance and preference, mobile phone technology has become the technology of choice because of portability, high computing speed and power [13].

ICT in agriculture most often include collaboration of two or more technologies. Agro-Climate is a term used to refer to decision support tools in agriculture that incorporates climate information in crop management. Research has already shown that there is great

potential in combining different ICTs for agriculture to provide smart technology in an effort to improve small-scale farmers' production [5]. In pursuit of this smart technology, artificial intelligence (AI) and the creation of intelligent systems has become the new and most popular topic in technology.

Artificial intelligence is the ability that is built into a computer system to be able to apply the intricate function that a human brain uses to learn and make decisions. Machine-learning and fuzzy logic have become the more widely used technological applications in the effort to produce intelligent systems. Fuzzy logic is a mathematical model in which logic depends on the degree of truth rather than a binary for where logic is either true or false. This kind of modelling is used in artificial intelligence, robotics, and business decision support systems. Machine learning (ML) makes use of neural networks, which is a data modelling tool that works by learning through numerous examples. According to Toseef *et al.*, [14], fuzzy logic can be applied for decision making for two reasons, “the rules are derived from expert knowledge that is described in natural language”, and “it handles the vagueness and uncertainty inherent in the problem”.

Research has suggested that collaborative technological solutions rather than individually employed techniques would yield much more desirable outcomes. Therefore, the aim of this investigation was to enable the possibility to determine the possible development of a portable, cost effective, accessible, and sustainable decision support tool for small-scale farmers. It seeks to form a symphony of multiple ICT disciplines, integrating indigenous knowledge as well as mobile phone technology to forge a smart Agro-climate decision support tool.

1.2 Problem statement

According to the state of food security and nutrition in the world report of 2021 by the Food and Agriculture Organization of the United Nations (FAO), South Africa is one of the countries whose food security is under threat. South Africa has shown an increase of almost 2% in prevalence of severe food insecurity since 2018 [15]. Globally, a steady decline in agricultural crops has been recorded [16]. With agriculture being the key driver for food security and economic sustainability, more attention has been drawn to small-

scale farmers as possible contributors towards food security improvement. Hence, in reaching various cropping decisions, small-scale rural farmers in Pietermaritzburg, South Africa have used a prediction tool called “the indigenous knowledge (IK)” for decades [11]. This has become the legacy from generation to generation. However, this method has become capricious and has had the most adverse consequence on productivity. The unreliability of these IK indicators is aggravated by the constant change in the climate due to global warming, unexpected droughts, and other known and unknown factors. On the other hand, this has also affected weather and climate predictions upon which scientific cropping decisions are based, making them inadequate tools. Because of these challenges, small-scale farmers in Pietermaritzburg have the most urgent need for reliable and cost-effective operational support in a form of synoptic information that aids towards actionable decisions in terms of day-to-day, medium-term, and long-term decisions. Indigenous knowledge alone is currently not sufficient owing to the instability and inconsistency of the natural environment. Therefore a collaboration of ICT technologies in agriculture provides a better opportunity for smarter technology [17]. If this informational gap is not filled, the small-scale rural farmer will continue to suffer financial losses and experience productivity disasters [12].

1.3 Objectives of the study

The overall aim of this study was to develop and evaluate the effectiveness of an integrated portable, cost effective, accessible, and sustainable smart Agro-climate decision support tool. The tool sought to form a symphony of multiple ICT disciplines that is wireless sensors and machine learning, integrated with indigenous knowledge as well as mobile phone technology to forge a smart Agro-climate decision support tool that works for small-scale farmers. This was guided by the following specific objectives.

- To collect and analyse data on the current indigenous and scientific approaches that small-scale farmers in Pietermaritzburg use in reaching various cropping decisions.
- Based on the results of the data analysis, to develop an integrated (indigenous knowledge, agro-climatic and edaphic conditions) intelligent agro-climate model using Artificial Intelligence, Machine Learning (ML) and Fuzzy logic in this case.

- To evaluate the extent to which the derived decision support model or algorithm will improve the level of accuracy of the cropping decision-making and farm productivity by the small-scale farmers.

1.4 Research methodology

A combination of research methodology approaches was used in this research and both qualitative and quantitative methods were employed. Case study research focusing on the small-scale farmers in the Swayimane area of Pietermaritzburg in the province of KwaZulu-Natal, in partnership with the University of KwaZulu-Natal's faculty of Agriculture, was carried out. Input data that constitutes climatic and edaphic conditions was collected. Adaptation of the framework for the ITIKI [9] and the Intelligent mobile system model [14] was leveraged in the design of the proposed system. The system made use of inputs from observed IK indicators according to farming activities such planting, harvesting, and fertigation. These inputs are then run through the fuzzy inference system. The choice of basing the system on ITIKI framework is because the bulk of a farmer's major decisions is related to rainfall and for which ITIKI already predicts. A lookup from crop science database provided the output information that allowed for Artificial Intelligence models and algorithms to be employed.

Data from the farmers was collected through a series of focus group discussions held with the farmers. This helped in determining the kind of IK that these farmers make use of in their cropping decisions. Within these focus group discussions, questionnaires and interviews were used to further explore issues which were initially not clearly understood. In the original plan for this study, sensors were supposed to be installed in the field for the purpose of collecting of edaphic factors. However, due to the timeline of the research and the limitations that came with the restriction resulting from the Covid-19 pandemic, the information that would have been collected through sensors was collected from crop science database available from the Department of Agriculture. This information assisted in creating the necessary data files for simulation of the Artificial Intelligence. This method of data collection is called feature construction or augmentation [18].

Finally, framework development was applied for the creation of an extended version of ITIKI framework - the ITIKI Plus. ITIKI Plus enhances ITIKI by incorporating cropping information to support farmers' tactical and routine decisions.

1.5 Contribution to the knowledge

The study reviews agricultural technologies that are in existence globally, and zeros in on what is available for Sub-Saharan Africa. It contributes towards bridging the technological divide that exists between the commercial agriculture and small-scale agriculture, by initiating collaborative solution-based technology. The study takes into consideration existing technologies, then evaluates and presents a comparison of these technologies to derive the best suited agricultural technology for small-scale farmers. The study takes advantage of the old (accessible, tried and tested) to recreate the new (a symphony of collaborative ICT technologies). This study also uses the ITIKI framework to include the Indigenous Knowledge as part of the solution.

1.6 Research publications from the study

Chapters in a book

Thothela N.P., Markus E.D., Masinde M., Abu-Mahfouz A.M. (2020). A Survey of Intelligent Agro-climate Decision Support Tool for Small-Scale Farmers: An Integration of Indigenous Knowledge, Mobile Phone Technology and Smart Sensors. In: Fong S., Dey N., Joshi A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 154. Springer, Singapore. https://doi.org/10.1007/978-981-15-8354-4_71

Conference paper

Thothela N.P. and Masinde M., ITIKI Plus: A Mobile Based Application for Integrating Indigenous Knowledge and Scientific Agro-Climatic Decision Support for Africa's Small-Scale Farmers. In 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), pp. 303-309. IEEE, 2019.

Submitted

Thothela N.P., Markus E.D., Masinde M., Abu-Mahfouz A.M. (2022) An Intelligent Agro-Climate Decision Support System for Small-scale farmers (submitted).

1.7 Limitations of the study

This study was limited to modelling a decision support tool for small-scale farmers only. The study focused on using the agro-climate, indigenous knowledge, and edaphic conditions of a particular area in the province of KwaZulu-Natal, South Africa to model and simulate the Artificial Intelligence using machine learning. The study focused on crop management agriculture and considered only crops that are suitable for the selected area. The deployment and testing out the solution on mobile phones are beyond the scope of the study.

1.8 Outline of the dissertation

Chapter 1 offers an introduction to the study, and it provides the background and lays the groundwork for the study. It outlines the motivation and the reasoning behind this study. It further sheds light on the problem statement.

Chapter 2 provides a literature review that supports the motivation for the study and outlines the work that has already been done in this field of study.

Chapter 3 presents the research methodology of the development of the solution and the discussion of the study area.

Chapter 4 provides a discussion of the analysis of the data used and the simulations of the ML algorithms.

Chapter 5 is a discussion of the results of the simulations of the ML algorithms.

Chapter 6 provides the conclusions and recommendations for further studies and considerations.

Chapter 2: Literature Review

2.1 Introduction

In this chapter, a thorough evaluation of the current situation in Sub-Saharan Africa agricultural economic sector is presented. The current challenges and possible solutions as researched are presented. The technologies that have been investigated and employed severally and in collaboration in this economic sector are reviewed and explained. In a quest to determine the path to developing and evaluating an intelligent agro-climate decision support tool for small-scale farmers, this chapter will explore the questions: What is the main challenge faced by small-scale farmers? What possible solutions and technologies are already available? Why the need for the current proposed innovation?

This chapter has been divided into the following topics:

- Mobile phone technology in agriculture
- Integration of Indigenous knowledge and ICTs within small-scale farming in Africa
- Decision support systems for small-scale farmers
- Application of Artificial Intelligence in cropping decision support
- Discussion and comparison of considered technologies
- Conclusion

2.2 Overview of Mobile phone technology

In 1947 when the transmitter was invented, Douglas H Ring, an engineer at the Bell Labs, sketched out the rough design for a standard cellular phone network. It would take 40 years for technology to catch up with that vision. Although Martin Cooper of Motorola is considered the inventor of the hand-held cellular telephone, he used technology developed by Bell Labs engineers. Since the very first invention, mobile phone technology evolved five generations, and transformed the way we live. The first generation of mobile phone technology basically used analogue radio signals. It was

gradually replaced by the second generation which incorporated digital networks, and the first introduction of Short Message Service (SMS), and Multimedia Messaging Service (MMS). The third generation introduced multimedia application, and internet application with high connectivity speed. The fourth generation improved on the internet connection speed and the fifth generation (5G) is expected to solve most major smart phone issues and the security over the internet [19]. Modern applications utilising the 5G include video conferencing and real-time video application. Others include virtual reality and augmented reality, which are becoming more popular owing to Artificial Intelligence.

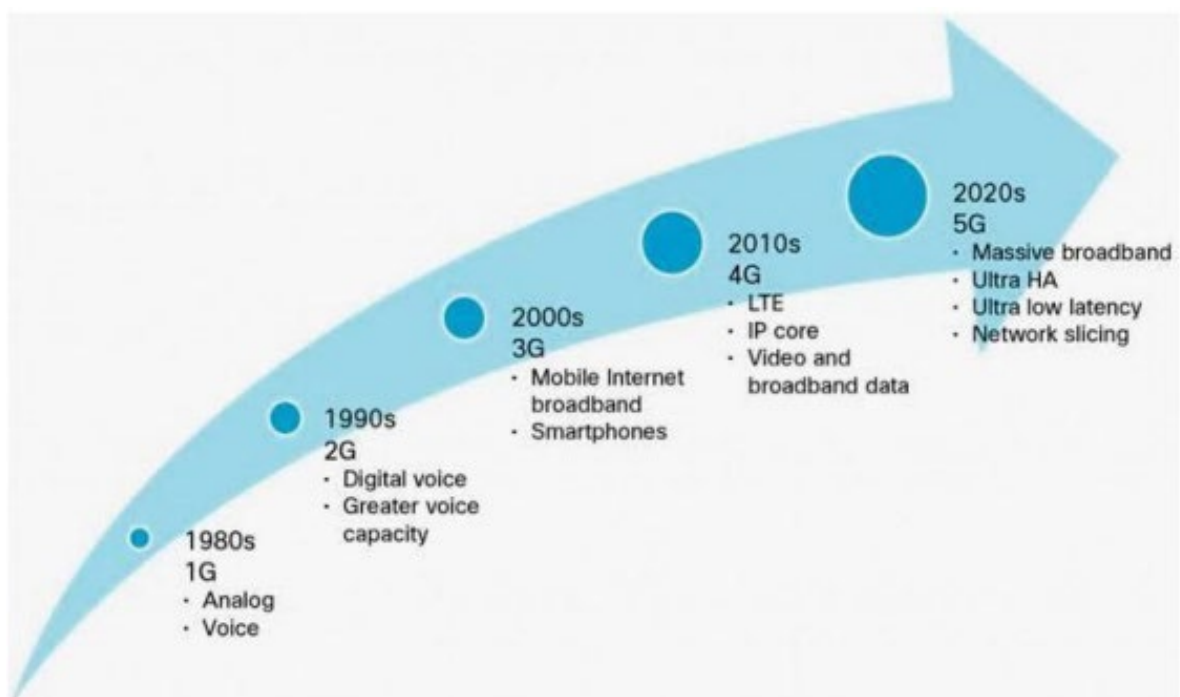


Figure 2. 1 Evolution of mobile phone technology [19]

2.2.1 Mobile technology in Sub-Saharan Africa

In terms of ICT indicators, though Africa has made a significant improvement in terms of internet use, it is still far behind counterparts from other regions [7]. In Africa, where the footprint of mobile phone stands higher than 63% on sales [8], idiosyncrasies is towards mobile broadband as an internet connectivity of choice because the cost of mobile broadband is fifty percent lower than that of fixed broadband. According to

Masinde *et al.*, [8], mobile phone technology has advanced to the point that phones can now compete with personal computers from less than a decade ago; the devices (particularly smart phones) are no longer just phones for making and receiving calls. The mobile phone added advantage is that it can act as both input and output device, owing to its highly technologically developed user interface.

The mobile phone technology has become a gigantic form of communication worldwide, and mobile phone networks have gained an explosive growth in the past [12]. According to Kabbiri *et al.*, [12], mobile phone technology in Sub-Saharan Africa has enhanced dissemination and retrieval of information for people, particularly in the rural communities. The author discovered that in Africa, mobile phones overtook the number of landlines. Mobile phone ownership has increased faster in Africa than any place in the globe since the 2000s [20]. The author points to the fact that 83% of the population in Africa currently has mobile subscription.

Mobile phone technology has also had an impact on the decrease of certain costs, resulting in improved functioning of certain sectors including agriculture, and there has been great improvement in the way business is conducted. Sub-Saharan Africa continues to lead the world in adoption of mobile money services [12], [21], and in Sub-Saharan Africa mobile industry continues to experience exponential growth reaching 774 million subscribers in 2018, and expected to reach 1004 million by 2025. Mobile phones are the most easily accessible and cost-effective technology available in South Africa. Africa has a far greater mobile phone penetration rate than it does for computers [11]. It is apparent therefore that it would be easy for any small-scale farmer operating anywhere to acquire one.

Mobile communication technology has become the most quotidian manner of transmitting data and services in a world that is constantly evolving. This drastic change gave rise to mobile applications [21]. The author suggests that mobile applications for agricultural and rural development possesses potential for advancement of rural development and can provide an affordable medium of information access to numerous people who were previously disadvantaged. Smartphone shipments into Africa grew by 13.2% in the second quarter of 2021. With the growth in smartphone use, there is growth

for mobile application use. Mobile phone technology seems to be bridging Africa's pronounced gap between the rural demographic group with restricted access and the more urban demographic group with access to the latest ICT [12], owing to the fact that in comparison to other ICT mobile phone technology is much more affordable. Studies have also proved that if mobile phones were exploited properly, they could improve the influence of Sub-Saharan African farmers in the value chain [12].

It is of utmost importance to explore intentions of small-scale farmers to adopt technological inventions that affect them [22]. The study showed that expectancy of performance and effort, price value and trust proved critical to the adoption of such. In [12], [23] the author eluded to the fact that the farmers would have to be educated to realise the full potential of the mobile phone technology that is being offered in order to gain their buy-in for the purpose of the adoption or acceptance of the ICT in their space. According to Kabbiri *et al.*, [12], studies have shown how technology and digitisation impacts rural communities and that it can aid curb hindrances such as information access challenges for small-scale farmers.

2.2.2 Mobile technology in Agriculture

According to Kabbiri *et al.*, [12], one of the most neglected global challenges that Africa faces is digital divide. However, the author alludes to the fact that this colossal challenge is being restrained by the adoption of mobile phones. In agriculture, technology based decision support tools are typically software applications that are commonly based on models that describe different processes in farming [24], [21]. Mobile phone technology has a lot of potential in terms of progressing development [21], but many technologies in agriculture, including those that are climate-smart are not being maximised because of low rate of adoption by small-scale farmers [1], [12]. Users of mobile technology in agriculture are also diverse and include several stakeholders. Mobile phone technology applications generally provide crucial information for rural small-scale farmers. DrumNet is one of the mobile applications used by Kenyan farmers, which has been instrumental in increasing their income by a third [25]. This service links the small-scale farmers to commercial banks and retail providers of farm inputs.

2.3 Integration of Indigenous Knowledge and ICTs in Small-scale Farming in Africa

2.3.1 Indigenous knowledge

The indigenous knowledge (IK), which seems to be challenged currently by the ever-changing climate, is the local based (undocumented) knowledge that the small-scale farmers have employed for many years as a prediction tool and decision support management tool. It is somehow a legacy for the community and passed down from generation to generation [11]. Indigenous knowledge is normally based on the cultural ties to the natural environment, and the interpretation of the behaviour of nature around, including insects, animals, and birds for example. This can include predictive systems such as those for rain or drought, ploughing and planting season, harvest success, among many others and is based on environmental circumstances and events. These events can include, as an example, the yellowing of leaves on a crop, which will have a meaning attached to it, depending on the respective community. IK has been the basis of the decision supporting tool used by small-scale farmers. The idea is not to ‘confiscate’ the IK they are already familiar with, but to turn it into an enhanced decision supporting tool that is merged with modern science and technology.

2.3.2 ICTs in Small-scale Agriculture

Information and Communication Technologies (ICT) in agriculture encompasses a plethora of technologies already in service [26], such as Global Positioning System (GPS), Geographic Information Systems (GIS), robotic systems, and many more. In South Africa, the agricultural industry is fractionated in two sectors. These include the large-scale commercial farmers, medium-scale or emerging farmers, and the small-scale rural farmers. In Sub-Saharan Africa however, analysis show a third fracture from small-scale farmers to medium-scale farmers which are differentiated according to size of land [27]. According to Jayne et al., the conclusion was that there is not much differentiation between the two, though medium-scale farmers are the source of dynamism, technical change, and commercialisation of African agriculture [27]. The differentiation between the small and large was defined by the Department of Agriculture. According to Kirsten

et al., the Department of Agriculture in their policy describe a small-scale farmer as one whose operational scale is too small for the attraction of the need for service providers for his/her operational productivity improvement [6]. The technology, mechanisation, and machinery available to support daily operations is in fact what separates the two, and the size has nothing to do with establishment of the difference. While state of the art technology and sophisticated machinery is at the disposal of the commercial farmer for the support of daily running, the small-scale rural farmer will support his/her daily operations through intense labour, relying on their own strength and working with their hands. Hiring or buying equipment is an unaffordable luxury for the small-scale farmers [28].

Yield forecasting seem to be playing a large role in agriculture. As already mentioned, a lot of research and resources have been put into ICT in agriculture because it is an important part of the economy and forecasting of crop yield is of utmost importance for food security [29]. Numerous research has been done in the quest to find the best forecasting method for crop yield [29]–[32]. In a study by Meroni et al., it was mentioned that climate data and remote sensor data have become monumental in crop yield forecasting and learning models for ML and deep learning have shown to be effective in a variety of data-driven applications, such as yield estimation. These models however require large datasets for accurate forecasting and a range of conditions.

Another aspect of crop agriculture management that has received much attention in terms of research is pest control or crop disease detection. This is another challenge that small-scale farmers particularly in Sub-Saharan Africa have to deal with [33]. In this study by Nyabako et al., prediction of an infestation of the notorious pest that normally attacks stored maize was researched using machine learning. In a study by Li et al., deep learning was used to determine plant health [34]. According to the author, early detection of plant disease can assist in curbing food insecurity. Crop disease detection is another aspect of agricultural crop management that is normally done manually by farmers inspecting the crops directly. This can be time consuming and requires extensive experience, otherwise misjudgements and misidentification can result [34]. Crop pest infestation is an enemy to good crop yield, and thus in turn threatens food security [35].

A review of machine learning in agriculture done by Liakos et al., in which different kinds of machine learning algorithms that have been utilised in agriculture were evaluated according to the areas of agriculture in which they were employed [36]. In this review, the works examined were divided into two categories: (1) crop management, which included applications such as yield prediction, disease detection, weed detection, crop quality, and species recognition; and (2) livestock management, which included applications such as animal identification and disease detection. The author also affirms the significance of yield prediction in crop management agriculture and highlights the role that ICT plays in agriculture in a series of projects that are mentioned. In the review, machine learning was used for different approaches in crop management which accounted for 61% of all the work reviewed [36].

2.3.3 Integration of indigenous knowledge

Potential in collaboration of ICT to provide smart innovation for agriculture was uncovered in a study by Amarnath *et al.* [5]. The study proved that “the greatest impact of ICT can be expected in areas where farmers’ actions are currently the primary limitation to their production”. The author emphasises that advancement in a single aspect of agricultural crop farming without taking into consideration all other relevant aspects such as soil tillage, geographic location, tolerance of environmental stress, soil type, rainfall, soil fertigation, pests and disease control and nutrition can be economically inefficient. The study concluded that it was critical to scale up on agricultural ICT services for small-scale farmers, but also that these services should be delivered at tailored local scale, otherwise they would not be relevant and would not be able to support the small-scale farmers’ decision management.

A study reported in [37] highlighted the need for integration of indigenous knowledge in agricultural ICT development. According to the author, “farmers greatly value local experiential knowledge as they see it as having practical, personal and local relevance”. It is therefore important to deliver information to small-scale farmers in a customised format that is relevant to them. In actual fact, studies have uncovered that more than 80% of farmers in countries such as Kenya, Zambia and Zimbabwe in the African continent are still dependent on their indigenous knowledge as a form of forecast on which their

daily agricultural management decisions are based [4]. In the study by [9], the author quotes, “the effectiveness of the forecast information depends strongly on the systems that distribute the information, the channels of distribution, the recipients’ models of understanding and judgment about the information sources, and the way in which the information is presented”.

There is a strong relationship that exists between IK and the scientific Decision Support Systems (DSS). This relationship has been accepted as rather complementary areas of expertise which formed the basis of countless relevant and appropriate technologies that have been inaugurated to date [11]. Researchers are also in one accord to the certitude that the integration between the two “can improve livelihoods”. According to the author, “In order to build sustainable strategies, it is therefore important to take into account of, and learn from what the local people already know and do”. As much as Indigenous knowledge and science may complement each other, they are luminously contrasting expertise. Considering that indigenous knowledge indicators have come under threat owing to the change in climate and global warming, it can be argued that science is needed to support the fading discipline.

Meteorological organisations in Africa have a colossal task with seasonal weather forecast generation. They are equipped with currently very few weather stations scattered over a vast area. This is an opportunity to introduce technologies that will help to bridge the gap and the indigenous knowledge of the local communities can contribute to the solution, giving them a sense of ownership of a system that is delivered. There is an opportunity that can be exploited through unclouded comprehension of indigenous knowledge and appropriate integration with science, particularly in the distribution of weather forecasts to farmers in rural settlements.

In pursuit of the solution to the question, “does incorporating indigenous knowledge into the drought prediction tool improve the resilience and relevance to the countries in Sub-Saharan Africa”, in a study by [9] a framework was developed for integration of mobile technology and indigenous knowledge for Africa. The idea is to enhance what the small-scale farmer also has, rather than replace it completely to curb resistance in adoption of the deliverable system. According to a study by [12], the conclusion was that “perceived

advantage and perceived usefulness influences mobile phone adoption negatively”. This was attributed by the fact that most small-scale farmers only used their mobile phones for normal communication and not as an assistive tool for their agricultural activities. Incorporating what they already know that is their indigenous knowledge, could play a role in encouraging adoption of the deliverable system. This also encourages multi-purpose use of their mobile phones.

2.4 Decision Support Systems for Small-scale Farmers

“Agricultural decision support tools are typically software applications, commonly based on models describing various biophysical processes in farming systems and the response to varying management practices” [24]. Decision management is at the heart of agricultural practices and operations, and decision management is in turn dependent upon or is supported by access to information such as climate forecast, soil fertigation, soil type, humidity and so on. Crop farmers always need to be cognizant of these factors to be able to make the right decision. According to [38], “while too little or too much rain can kill crops, the proper amount of rain leads to an ideal crop yield.” Hence, small-scale farmers need timeous forecast on weather and climate. In sub-Saharan Africa, it is a gargantuan task for small-scale farmers in the rural areas to obtain seasonal climate forecasts even though meteorological institutions constantly and regularly provide it, because of the manner in which the information is distributed [9].

A plethora of variable crop prediction models have been investigated to predict various aspects of crop agriculture [26], [30], [32], [38], [39]. It is lucid therefore, that crop prediction requires a synthesis of various considerations to come up with the best agricultural practices for maximum productivity. This study solicits a solution that amalgamates and brings into convergence various aspects of crop agricultural decision support tools into one system and simplifies decision support management for the small-scale farmer.

The concept of niche market/crop would be a luxury for the small-scale farmer, and therefore because they do not specialise, they often have a variety of crops planted in the same field. This is referred to as intercropping, and studies were conducted to determine

how to deal with the challenge of water tolerance in intercropping [40], and to determine what crops would thrive if paired in intercropping process [41]. According to [42], well managed and well balanced intercropping can improve the productivity when crops that have been paired benefit positively from one another in terms of nutrients and water which would be a shared source. Though intercropping has been used by small-scale farmers, it has not been exploited to full capacity due to lack of information. According to [42], “Farmers need an efficient, relevant, and accurate way to evaluate data for specific management decision”. This presents another opportunity for a solution that can be incorporated into the proposed system.

According to Toseef *et al.*, [14], one of the challenges that are faced by small-scale farmers in developing countries is undiagnosed crop diseases or inefficiency in the diagnosis of crop diseases. This study also found that some of the drawbacks in existing systems in agriculture generally include insufficient knowledge about the diseases and the farmers’ lack of education. According to Phelan *et al.*, [24] a well-designed agricultural decision support system equips farmers with a facile and rapid way of comparing multiple scenarios for management of crop production decisions. One example of such tools is CropARM, which assists farmers with management decision actions such as planting, by establishing risks taking climate into consideration have been designed [24].

According to Jahel *et al.*, agricultural statistics divulge inter-annual variations in the quantities of harvested crops that can be attributed to climate change, global warming, and economic uncertainty [43]. The variation in the climatic becomes difficult to keep track of it if it is not properly documented and causes environmental uncertainties. It is important to understand these climatic variations since they affect the agricultural decision the farmers have to make regarding their crop which in turn directly affects the quality and quantity of the harvest [43]. Agricultural farming is also impacted by economic circumstances, access to credit and markets [43]. With such a multitude of external factors affecting operations in farming, small-scale farmers find themselves having a difficult task of having to make strategic, operational, tactical, long-term and short-term decisions [43]. Without relevant information it becomes a daunting task for small-scale farmers to make informed decisions.

Agricultural farming is also impacted by economic circumstances, access to credit and markets [43], [16]. With such a multitude of external factors affecting operations in farming, small-scale farmers find themselves having a humongous task of having to make strategic, operational, tactical, long-term and short-term decisions [43]. Without relevant information, it becomes a daunting task for small-scale farmers to make informed decisions.

2.5 Application of Artificial Intelligence for Cropping Decision Support

2.5.1 Application of Machine learning for cropping decisions

Machine learning (ML) as a subsidiary of Artificial Intelligence (AI), has been described by Apro Software as “Agriculture’s new best friend” for a number of reasons [44]. The applications thereof are already employed in crop management, soil management, water management, as well as livestock management in agriculture. In a study by [31], the author suggested that Artificial Neural Networks (ANN) could be used in collaboration with other AI methods or statistical techniques to avoid limitations possessed by ANN. ML has the capability to independently solve large non-linear problems using datasets from several sources as an advantage [45]. In the precision agricultural study by [45], the author concluded that ML, in combination with remote sensing, will continue to provide comprehensive solution for better estimation and decision support. ML offers a strong and versatile platform for incorporating expert knowledge as well as data-driven decision-making processes. Without human interaction, ML promotes improved decision-making and substantiated decisions in real-world scenarios. In a Machine Learning process, a ML model goes through a learning process by making use of datasets. The ML model or algorithm that is chosen is highly dependent on the type of problem that needs to be solved. A performance metric is used to assess a model's performance. At the end of the training, the model can then be used to make predictions, classify or cluster based on the knowledge obtained from the training [36]. In a review done by this article [36], the most commonly used model in agricultural crop management is ANN.

2.5.2 Fuzzy Inference System

Since the introduction of the Internet of Things (IoT), the need for intelligent systems has grown in popularity. Artificial intelligence is the ability that is built into a computer system to be able to apply the intricate function that a human brain uses to learn and make decisions. Machine-learning and fuzzy logic have become the more widely used technological applications in the effort to produce intelligent systems. Fuzzy logic is a mathematical model in which logic depends on the degree of truth rather than a binary for where logic is either true or false. This kind of modelling is used in artificial intelligence, robotics, and business decision support systems. According to Toseef *et al.*, [14], fuzzy logic can be applied for decision making for two reasons, “the rules are derived from expert knowledge that is described in natural language”, and “it handles the vagueness and uncertainly inherent in the problem”.

Fuzzy inference system can be used in this regard because it would take as input the information entered by the small-scale farmers as described according to their understanding. The information the farmer enters may be “fuzzy” or vague. This is because fuzzy logic is a multi-valued mathematical model that is able to work on the degree of truth rather than either true or false. Fuzzy logic may be employed in image processing, decision making and other artificial intelligence kinds of systems. Fuzzy logic magic may also be used over Android mobile. An intelligent framework was developed in the study by [14] for diagnosis of crop diseases. A Fuzzy Inference System (FIS), which is a rule-based system that makes use of fuzzy logic instead of normal Boolean logic, is exploited in this study. FIS provides an intelligent engine for synthesising vague and uncertain information provided as part of the decision-making process.

The FIS process has three different stages as explained in [14], and as depicted in Figure 2.2. First, there is an input from the environment that must be fuzzified. This input passes through the inference engine where the fuzzy rules are applied using the fuzzy logic. A mapping according to the degree of truth is used to get a fuzzy value (z^*). This value is used in one of two methods of defuzzification, Centroid which uses the centre of area or Middle/Mean of maximum (mom) which is self-explanatory. The following formulas (equation 1 and 2) applies:

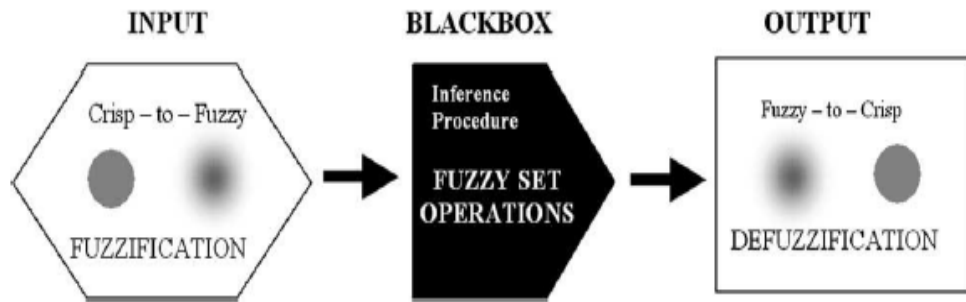


Figure 2. 2 Stages in fuzzy modelling [46]

Centroid
$$z^* = \int \mu(x) x dx / \int \mu(x) dx \quad (1)$$

Mom
$$z^* = a + b/2 \quad (2)$$

The z^* is the distance of every relevant maximum value of the function to the centroid, and \int is the algebraic integration of the sum of the. The value of z^* is calculated using the average sum in the Centroid formula and the mean of maximum in the MOM formula.

2.5.3 Application of Fuzzy Inference System to cropping decisions

In the study by [14], an intelligent expert system for diagnosis of crop diseases was developed. The architecture for this system used fuzzy logic as a backend decision engine and was designed to work with Android mobile technology employing jFuzzylite library. Figure 2.3 depicts the architecture of the intelligent mobile application for crop disease diagnosis according to Toeseef *et al.*, [14]. According to the author, fuzzy logic was used for decision support in this frame for two main reasons. The first reason is that the rules are derived from expert knowledge described in simple natural language and fuzzy logic is “a powerful knowledge representation mechanism for linguistic knowledge”. The second reason is that it can handle uncertainty of the input thereby providing the farmer with the opportunity to be as intuitive as possible in their input guesses. This also explains why fuzzy logic was employed [9] in the development of the ITIKI framework as shown in Figure 2.4.

2.6 Discussion and comparison of considered technologies

The ITIKI system did well in incorporating the indigenous knowledge, Artificial Neural Networks and the sensor technology to assist small-scale farmers with drought prediction [9]. This is a crucial solution considering that small-scale farmers operate in rainfed agriculture. Rainfall is only one part involved in agricultural decision making. This prediction tool has been instrumental in the improvement of the farming decisions in Sub-Saharan Africa. The challenge that the system has is that it does not provide the farmers with holistic agricultural information required to make strategic, tactical, and operational management, such as decisions on fertilizer options, choice of crop for different conditions provided, agricultural drought, and the intercropping potential to exploit water resilient crops for maximum benefit.

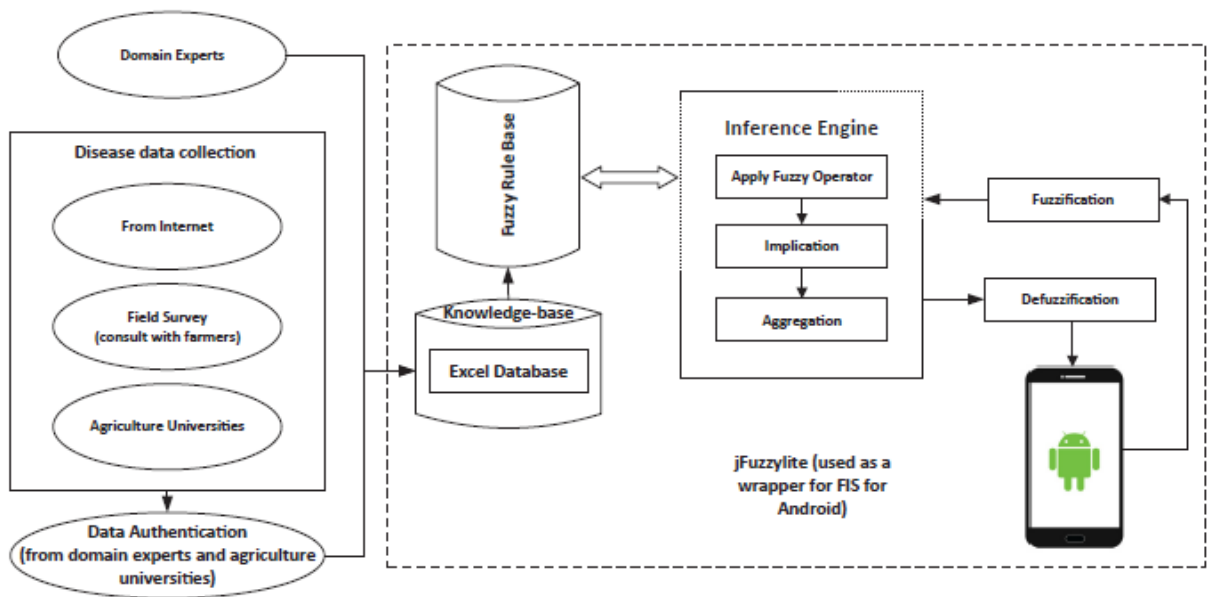


Figure 2. 3 Intelligent mobile application for crop disease diagnosis [14]

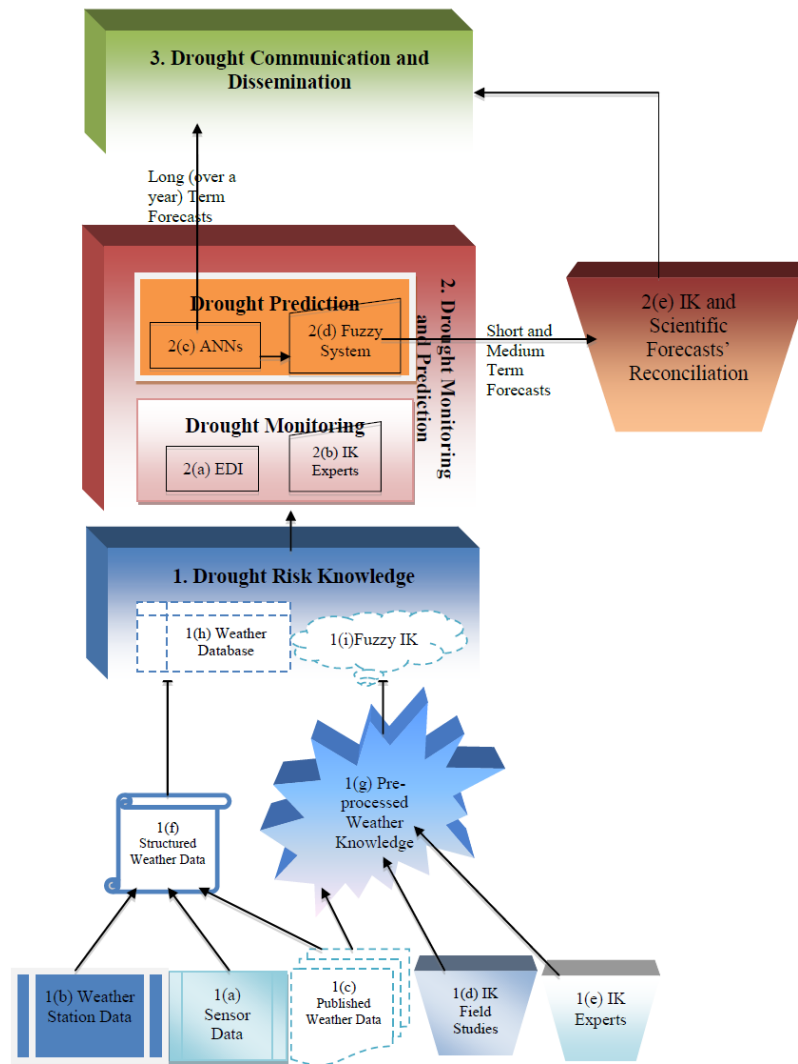


Figure 2. 4 ITIKI: Integrated Drought Early Warning System Framework [47]

The advantage of the ICT solution in the intelligent mobile application system applied in the study by Niebel et al., [14] is that it allows for uncertainty of input to go through a fuzzy inference engine to provide a prediction on crop diseases. The system also employs the use of mobile phone technology, which is a kind of technology that is more accessible to the small-scale farming technology in Sub-Saharan Africa. The challenge of the system by Niebel et al is that it does not allow for expert authentication of the output at runtime. The system does not cater for a more encompassing solution towards agricultural challenges faced by small-scale farmers in rural societies, making it a single target problem solving solution.

While SIMAGRI, also an Agro-climate system was aimed at assisting strategic and tactical decision in the production of crops, it uses historical weather to predict climate crop yield by running a simulation of “what if” scenarios [17]. The advantage of this system is that it takes a lot more management disciplines into consideration, such as planting dates, fertilizer type and environmental conditions, in addition, it takes user input. The disadvantage however is that the system simulates yield predictions in a graphical manner, which is not what small-scale farmers in Sub-Saharan Africa need.

CropARM is one other tool that was developed in the study by [24] in order to assist users to establish a framework of risk. It used the APSIM model to simulate scenarios using climate conditions and management actions. Though this system is available online and assists in management decisions that generated awfully close to predicted yields, it is not accessible to small-scale farmers in rural areas. In the study by [43], one of the limitations was determining the influential factors to cropping plans, meaning that there was no clear indication of the reasoning behind tactical decision. The author alluded to the fact that the small-scale communities based their choices on what the society in general was doing. This clearly shows that without enough resources and information, small-scale farmers’ agricultural decisions could clearly be hampered.

Table 2.1 highlights advantages and disadvantages of four relevant (to the current study) agricultural solutions that already exist. The solutions discussed had phenomenally successful results in their respective investigations, but this study emphasises the target users of the envisaged system under scrutiny. The disadvantages are therefore highlighted in the comparison table. For instance, SIMAGRI is a highly specialised system that supports multiple disciplines of agricultural management using historic data, making it difficult to enable the use for decision management for small-scale rural farmers. On the other hand, although CropARM establishes a framework of risk in agriculture by merging climate scenarios with agricultural management actions, it provides a complicated interface for output or results which would make it difficult for a small-scale farmer to use. This makes both solutions highly inaccessible to small-scale farmers due to the lower literacy levels among the mostly rural-based small-scale farmers. It is apparent therefore that these solutions are not suited for target user in this case.

The intelligent mobile application for diagnosis of crop diseases is only utilitarian in cases where crop health is in question, which is passable given the disaster that small-scale farmers experience when crop diseases go undiagnosed. This brings us to the question, ‘what about cases where there is no crop health scare’. How will this tool be of any use to a small-scale farmer? The answer is that it would be impractical. This disqualifies the solution for small-scale farmers because it is inaccessible to them. At this stage, ITIKI was found to be the only solution that is focused mainly on assisting small-scale farmers. It incorporates the small-scale farmer’s indigenous knowledge in a drought prediction tool, allows for vague user input, and was developed for small-scale farmers, making it a more user-friendly tool. Though this is only a drought prediction tool which is a shortcoming documented in the paper by Masinde et al., [10], it provides a foundation for a decision support tool that is more encompassing of agricultural management and management actions. With the prescription of the foundation therefore an intelligent Agro-climate decision support tool can be instituted.

Table 2. 1 Comparison of existing technologies used

ICT solution	Technology used	Advantages	Disadvantages
SIMAGRI	DSSAT	<ul style="list-style-type: none"> • Supports multiple strategic and tactical agricultural decisions • Uses tercile seasonal climate and climatology for risk analysis 	<ul style="list-style-type: none"> • Can be exigent for non-experts to apply for the purpose of agricultural decision making
CropARM	APSIM	<ul style="list-style-type: none"> • Help establish a framework of risk by incorporating climate scenarios and management actions • Regular updated climatic streams 	<ul style="list-style-type: none"> • Inaccessible to small-scale rural farmers • Output method is not user friendly for the small-scale farmers

Intelligent Mobile App	Fuzzy logic Mobile phone technologies	<ul style="list-style-type: none"> • Able to interpret vague inputs • Uses mobile phones technology and android OS • Incorporates local language for the for the farmers 	<ul style="list-style-type: none"> • For diagnosis of crop diseases only • Does not take many other agricultural management components into account
ITIKI	Fuzzy logic Mobile phone technologies ANN	<ul style="list-style-type: none"> • Early warning prediction for droughts • Able to interpret vague inputs • Incorporates Indigenous knowledge and is user friendly for small-scale farmers • Built to be user-friendly for small-scale farmers 	<ul style="list-style-type: none"> • Does not take many agricultural management components into account, it is only a drought prediction tool
ITIKI-Plus	Machine learning Mobile phone Sensors	<ul style="list-style-type: none"> • Assesses suitability of crop upfront • Minimises the risk • Takes IK into consideration • Output method is designed for the small scale farmers 	<ul style="list-style-type: none"> • Improved yield still dependant on uncontrollable factors

The advantages and disadvantages of different technologies that have been researched in the past that are relevant to ICT in agriculture were highlighted and discussed in comparison with the system developed in this study. This study supports the statement

made in the study by Chlingaryan et al, [45] that states that the development of hybrid ML systems are likely to play a huge role in precision agriculture in the future. This study developed a hybrid ML system that can be utilised in crop selection preceding planting stage in precision agriculture. Contrary to the studies by [26], [30], [32], [38], [39] that focused mainly on prediction of crop yield for particular crops, this study focused on selecting the right crop for planting depending on the climatic and edaphic conditions. The aforementioned studies also focused on only climatic conditions to do the particular predictions, except for the study by [45] which also included an edaphic factor that is the nitrogen content. The findings of this study confirm that the major contributor in prediction is climate conditions, which is identical to other mentioned studies. This study is complementary to the ITIKI study because selecting the right crop for the right conditions is just as important as predicting whether there will be enough rainfall for a satisfactory yield. ITIKI predicts whether there will be enough rain, and this study predicts what crops will thrive in the given climatic and edaphic condition.

2.7 Conclusion

Research has shown that there is a decline in the production of crops in Sub-Saharan Africa, and other parts of the world. Crop agriculture has been the worst affected by global warming and climate change. Small-scale farmers are faced with many challenges in their daily operations including the fact that they do not have access to climatic information, inputs, finances, machinery, and education. It has been shown that a number of technologies have been developed as a result of substantial agricultural research, but none of them are accessible or suited for small-scale farming.

Like other economic domains, successful agriculture is extremely dependant on the nature of the operational, tactical, and strategic decisions made. Making the right decision at the right time, can influence the outcome of a harvest. The decline in crop production can therefore be restraint by giving small-scale farmers tools to support their short, medium, and long-term decision making. In this chapter, existing technological solutions have been evaluated exhaustively with the aim to finding the best fit for the Sub-Saharan small-scale farmer. The literature review revealed that none of them could satisfy the need of the small-scale farmers from focal case study - Pietermaritzburg in South Africa.

The literature review further uncovered that various research experiments have proven that the scientific models in decision support systems can have certain limitations as well [48], as much as they can be pivotal in agronomic modelling. This therefore provides an opportunity for exploring a technological innovation that encompasses Indigenous Knowledge, Artificial Intelligence, and sensor information. After intense comparison and consideration, it has become conspicuous that this solution would be the best suited for small-scale Sub-Saharan African farmers and can be delivered on a mobile phone.

Crop agricultural research has to a large extent focused on crop yield prediction, which is not surprising because the concern is on the diminishing quantity of crop production where food security is concerned. The prediction of crop yield in the studies that are mentioned in this chapter are focused on a particular crop at a time, and the factors that affect the prediction are mainly climatic. This study contradicts the rest in that it is mainly focusing on selecting the correct crop to plant under the given climatic and edaphic condition, taking into consideration that small-scale farmers do not focus on niche farming. The approach of this study is that it does not isolate the climatic factors but rather collates it with edaphic factors.

Chapter 3: Research Methodology

3.1 Introduction

This chapter presents the methodology that was followed in putting together the ultimate solution for the modelling of the Agro-climate decision support system (DSS) for small-scale farmers. It clarifies aspects such as the design of the research, the methods for data collection and the development of the framework for the DSS. The aim is to elucidate the systematic approach followed in this research. This also includes the conclusions drawn after consideration of certain aspects and facts of the study.

3.2 Research design

An inductive research approach was followed for this study. An inductive research approach refers to the kind of research approach that does not begin with a formulated hypothesis but is based on aims and objectives that must be achieved through the research process [49]. Based on the aims and objectives stated in Chapter 1, the following research designs steps were selected and applied in this work: machine learning modelling, framework development, case study and software prototyping. These research design steps are discussed in detail below.

3.2.1 Machine learning models

According to Robert J. Sternberg, intelligence is the ability to learn, understand and think in a logical manner from experience [50]. Machine learning (ML) is a part of artificial intelligence (AI) that assists machines to become intelligent. In a quest to achieve an intelligent Agro-climate DSS, ML was employed in this study. In order to successfully build a ML solution, a ML lifecycle is followed.

A ML lifecycle follows five steps: data collection, data modelling, data preparation (pre-processing), model training, model evaluating and deployment. Model evaluation also involves fine-tuning the parameters to improve performance, and finally making predictions. In this work, the machine learning processes were followed as shown in

Figure 3.1. However, the final phase (model deployment), was not explored in this study because it is beyond the scope of this research. The simulations of the ML process are presented in detail in the next chapter. The simulation on the data was performed in order to determine the best ML algorithm for classification of the selected crops.

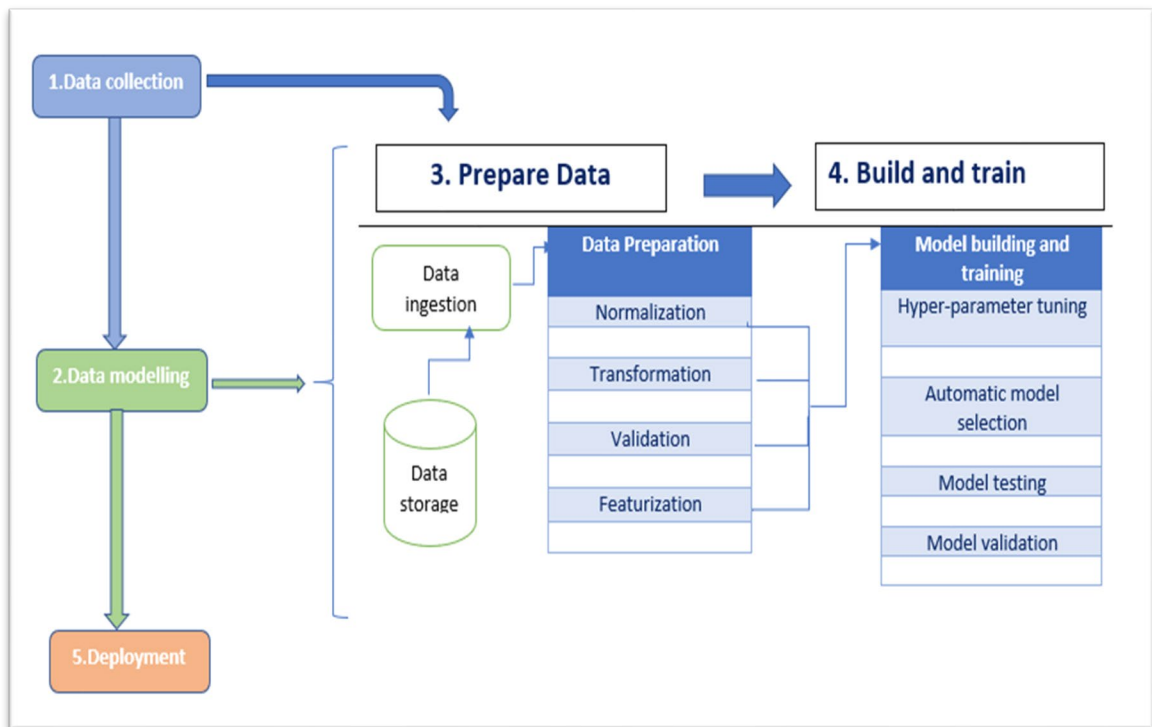


Figure 3. 1 Machine Learning process

3.2.2 Framework Development

As per objective 2 of this study, framework development approach was applied in defining the architecture of the integrated (indigenous knowledge, agro-climatic and edaphic conditions) intelligent agro-climate model. The development of the Intelligent Agro-climate Decision Support System followed the blueprint carved by the ITIKI Framework, which implemented the integration of Information and Communications Technologies (ICTs) with indigenous knowledge. Starting from the ITIKI Framework, the unique features of the proposed solution were used in extending ITIKI Framework to ITIKI-PLUS.

The ITIKI-PLUS Framework is based on cropping decisions made by small-scale farmers. As already determined in Chapter 1 the best cropping choice decision is dependent on the edaphic and climatic conditions, therefore data about climatic conditions and edaphic conditions affecting selected crops was collected. The data was run through a ML process employing different ML algorithms. Using supervised learning, ML classification was used to classify different crops according to the mentioned features. The data collected for the IKs was used to map and collaborate the cropping decisions made by the ML models. Mobile phone technology was employed for the dissemination of the cropping information.

3.2.3 Software prototyping

Software prototyping was used to build mobile-based application that demonstrates the implementation of the intelligent agro-climate model. The mobile application was built using Android studio. The application used an Android navigation pane as a home screen, with the details of the logged in user showing on the main screen. The application uses a static menu according to the user role of the logged in user. The design of the application used a consistent look and feel on all the menus to allow easy navigation.

The application allowed for a user registration prior to using the application, and a user account management such as password reset and so on. The application used Backendless for information storage and user account management. This prototype functionality was not tested in the field because of the Covid-19 lockdown. Figure 3.2 shows the user interface of the application. The images show that a logged in user is able to access two menus depicted on the left and the right. The middle picture shows reported IK which automatically tells the location of the reported IK using latitude and longitude. A reported IK is important for the purpose of this study, and will be discussed later as part of the intelligent agro-climate DSS model.

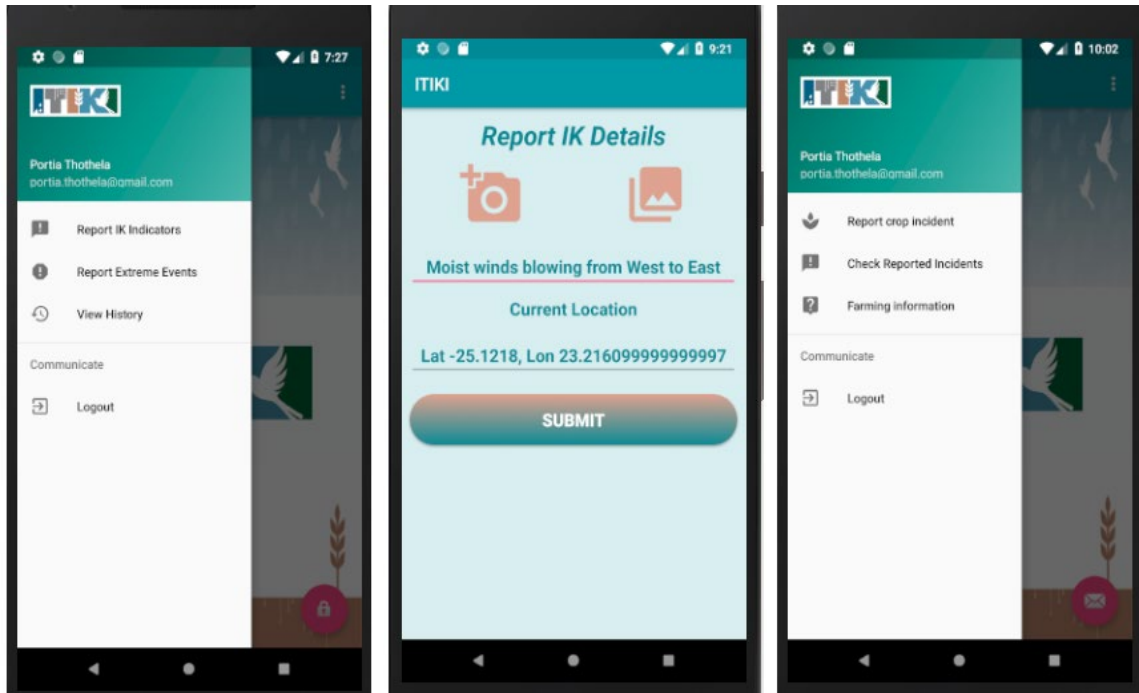


Figure 3. 2 The interface of the mobile application prototype developed

3.2.4 Case study

This research focused on the case study of Swayimane area, located in Pietermaritzburg in the KwaZulu-Natal (KZN) Province. Swayimane is a densely populated part of the uMshwathi ward 8 local municipality in the KwaZulu-Natal Midlands as shown in the map below. Figure 3.3 shows the municipalities around and surrounding the uMshwathi municipality in KZN province. uMshwathi local municipality is located in the district of uMgungundlovu. Swayimane is a suburb of Wartburg, with the coordinate's latitude $29^{\circ}25'50''S$ and longitude $30^{\circ}34'32''E$. Figure 3.4 shows a closer look at the uMshwathi municipality. The area marked in green represents the uMshwathi ward 8, and Figure 3.5 shows a closer look at ward 8 where Swayimane is located. Swayimane is situated 40km outside of Pietermaritzburg and is well known for its farming community. This small-scale farming community is mainly a crop farming community. Small-scale farming communities in this area mainly plant maize, beans, amadumbe, sweet potatoes and sugarcane. The staple food in Swayimane is maize.

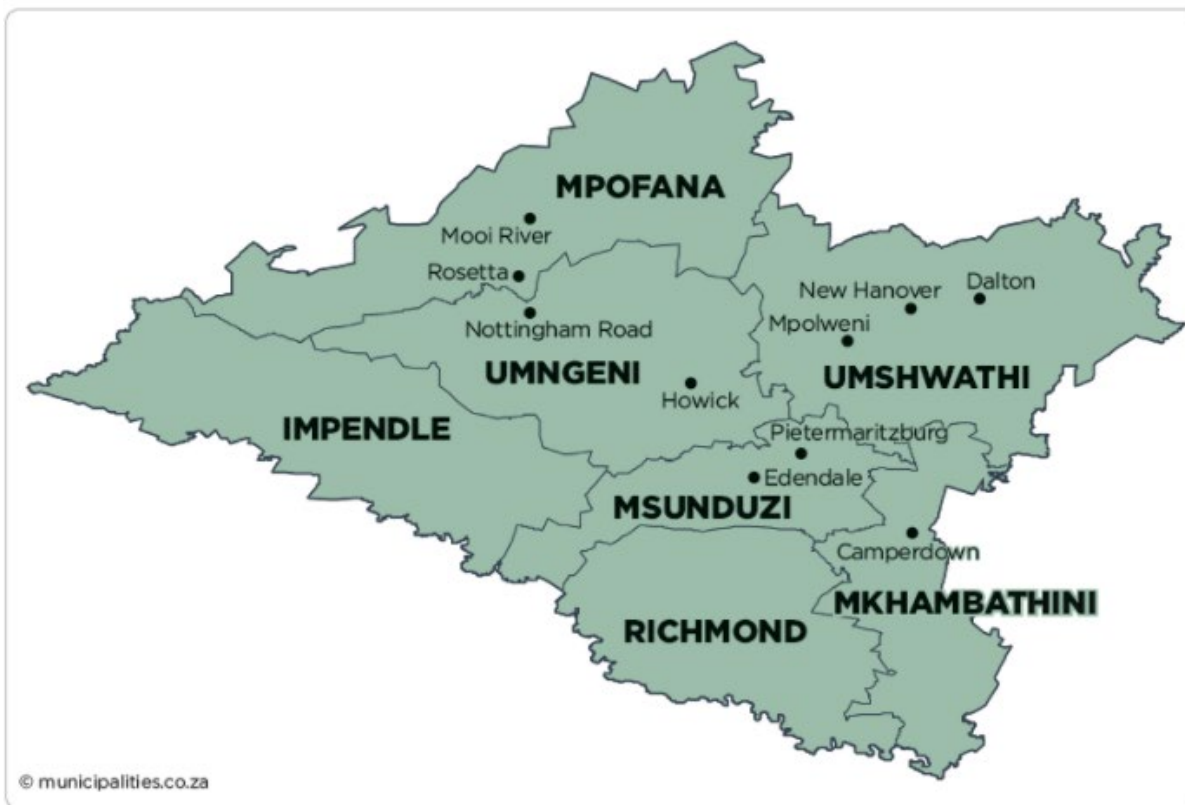


Figure 3. 3 Map of the municipalities in KZN midlands

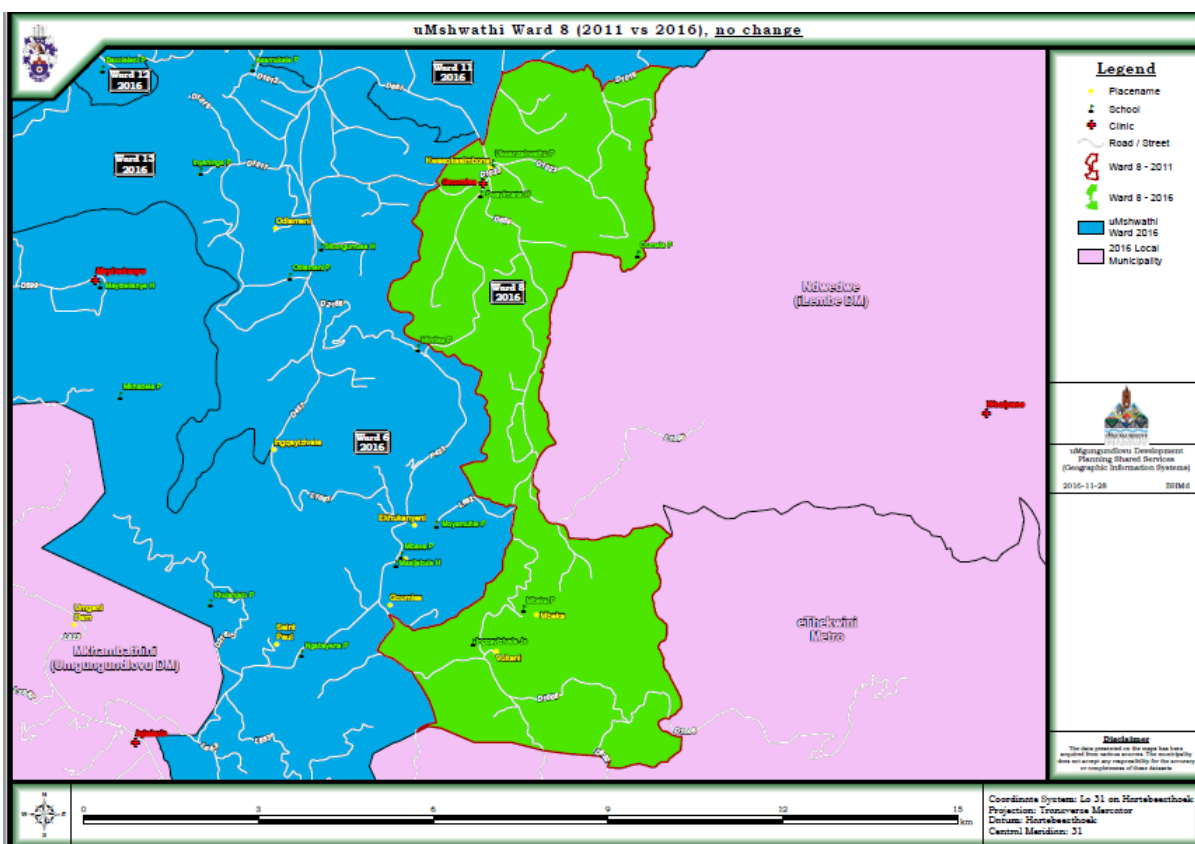


Figure 3. 4 Map of uMshwathi municipality ward 8

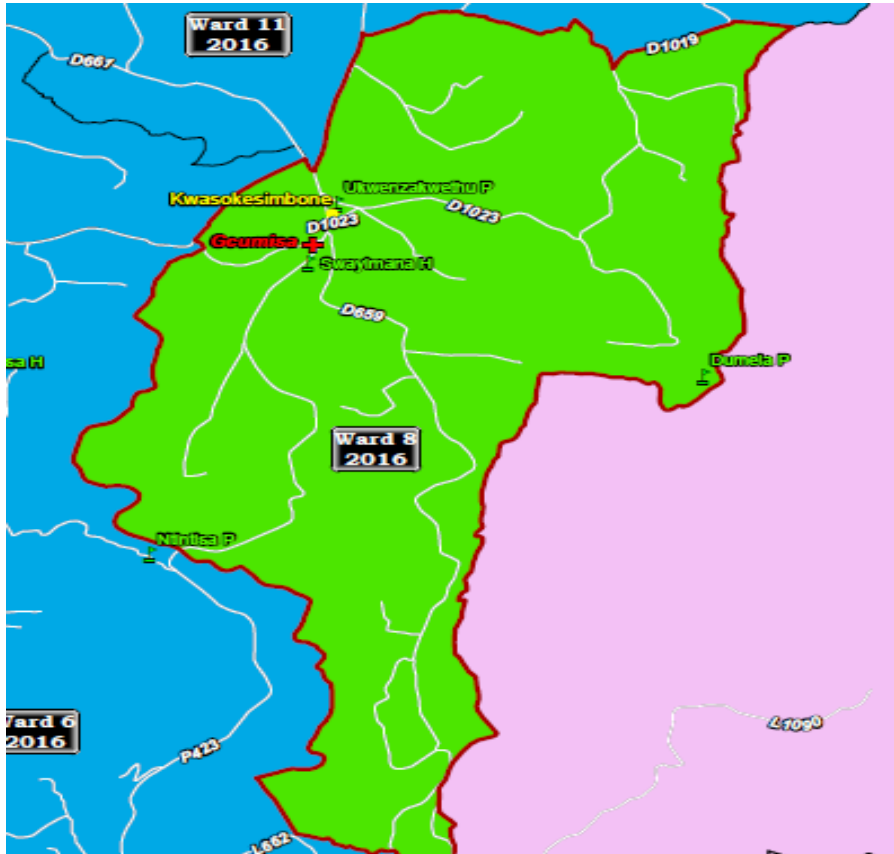


Figure 3. 5 Closer look at uMshwathi ward 8 (Location of Swayimane)

3.3 Data collection

Three sources of data were used: IK data collected using questionnaires, edaphic data and weather data sourced from the South African Weather Service. IK refers to the local knowledge used by the small-scale farmers of Swayimane for their decisions regarding daily operations, and edaphic data refers to factors that are related to the soil conditions that affect cropping decisions. Edaphic data is generally collected using suitable soil sensors, but for this study no sensors were deployed because due to the Covid-19 lockdown the appropriate time frames to order and install the equipment was missed. Weather data refers to the meteorological or climatic data.

3.3.1 Indigenous Knowledge Data

The collection of IK was done using questionnaires. As detailed in Appendix B, the questionnaire was sub-divided into 5 different sections. The first section was for the

demographic information, such as age, sex, the level of education and the level of experience as a farmer. The other sections of the questionnaire included the knowledge of weather forecasting, such as the type of weather forecasting used. IK before cultivation formed part of the questionnaire as well, as in the type of IK used for forecasting. It was important to determine the level of awareness of climate change and biodiversity as well, therefore another section of the questionnaire included questions to this effect. The last part included the indicators used to determine the offset of a season. *Table 3.1* is a representation of the IK collected for the offset of seasons for KwaZulu-Natal. Focus groups were used to provide more information on the details of the IK used particularly in the area.

Table 3.1 KZN Indigenous knowledge weather indicators

	SUMMER (December-February)	SUMMER (December-February)	WINTER (June-August)	SPRING (September-November)
Astronomical	<ul style="list-style-type: none"> • Full moon 	<ul style="list-style-type: none"> • The moon is small in size • Full moon 	Half or small moon	<ul style="list-style-type: none"> • Half moon
Meteorological (Knowledge of the Seasons)	<ul style="list-style-type: none"> • Very hot weather • High temperature during the day and night 		<ul style="list-style-type: none"> • Cold weather 	<ul style="list-style-type: none"> • It rains • Presence of thunderstorm and lightening • No rain sometimes. • Stars in night sky when it's going to be sunny
Behaviour of Birds	Magwababa and Inkonjane birds flock in before the rainfall.		Onogolantethe bird is flying looking for snakes and earthworms to eat.	Noisy migratory Birds are flocking in during spring e.g. The yellow Phezukomkhono bird.
Behaviour of Insects	Ntuthwane (Ants) are present, indicating the commencement of summer	<ul style="list-style-type: none"> • Insects are decreasing in autumn. • Sighting of Inyosi bees which are present in autumn 	<ul style="list-style-type: none"> • Insects are absent in winter. • Ants are hiding • No ants visible 	<ul style="list-style-type: none"> • Inyosi Bees are hiding in their hives. • Little insects

		indicate the autumn season.		
Behaviour of animals	<ul style="list-style-type: none"> • The animals are beautiful in summer, active and shinning fur. • Sighting of Ingxanga Frog in the undergrowth as the army of frogs move between the land and the watery homes. Laying their eggs – frogspawn – in any source of water they can find. • Cattles are gaining fat. • Most animals are getting fat. 	Cows are fat	<ul style="list-style-type: none"> • Little trace of bush animals, because it's cold, little activities. • The animals are thin • Thin, lost some weight 	The animals are average and lost weight
Flower, leaves and fruit production by some Trees.	<ul style="list-style-type: none"> • Flowering of some trees - Mviti Tree, Peach trees, Amapetjies, tshisi. • In summer flowers are blooming 	Some plants leaves are withering and decaying e.g., Mviti tree	Withering of leaves of some trees	<ul style="list-style-type: none"> • Blooming of some trees e.g Guava • Flowering of trees and plants e.g., Wattle, Wiki-jolo, Umphenjane.
Myths and religious beliefs				

This research sought the local understanding of the factors affecting cropping decisions before cultivation and planting for this group of small-scale farmers. The feedback in terms of the usage of the IK, the kind of IK used, and the ease of use was important. The likelihood of the adoption and acceptance of the DSS was a crucial part of this data collection since the solution would be tailor made for them. The other important aspects of this data collection were the understanding of the demographics in this community, the level of education, and the use of mobile phones.

3.3.2 Edaphic Data

Production guidelines for vegetables for KwaZulu-Natal as scientifically determined by the Department of Agriculture, Land reform and Rural Development and published in their website contained edaphic conditions affecting each crop documented in a form a brochure. To collect this data, brochures were obtained and analysed; each brochure contained requirements for the soil type, soil preparation, soil acidity, rainfall requirements, amounts of Nitrogen, Phosphorus and Potassium, weed control, pests, and so on. The edaphic conditions collected from these brochures were Nitrogen, Phosphorus, Potassium, rainfall, humidity, and pH. These were supposed to be collected using sensors, but due to the Covid-19 pandemic lockdown, the sensors were not installed. The brochures provided the minimum and maximum values of the selected features. Only five brochures were selected for this study and the information is summarised in Table 3.2.

Table 3. 2 Edaphic and climatic conditions affecting crop selection¹

Nitrogen (N)	Phosphorus (P)	Potassium (K)	Temperature	Humidity	Soil pH	Rainfall	Crop label
100 - 120	75 - 150	100 - 200	20 - 24	80 - 90	6,0 - 7,0	21 - 38	Tomatoes
150 - 200	11 - 28	75 - 240	18 - 38	90 - 95	5,0 - 5,4	460 - 910	Potatoes
10 - 40	150 - 250	100 - 200	20 - 25	50 - 70	6,0 - 7,0	25 - 600	Peppers

¹<https://www.dalrrd.gov.za/Branches/Agricultural-Production-Health-Food-Safety/Plant-Production/Production-Guidelines/Pguidelinesarchive>

60 - 100	35 - 60	15 - 25	19 - 23	55 - 75	4,5 - 7,0	350 - 450	Maize
40 - 60	55 - 80	100 - 150	18 - 24	18 - 25	5,8 - 6,5	400 - 650	Beans

The selection was informed by the KwaZulu-Natal planting calendar. The calendar is published by seeds for Africa for every province in South Africa. The mentioned planting calendar for the area was acquired from the website of DALRRD [51]. The selected crops are the ones that are typically planted by the Pietermaritzburg small-scale farmers. The soil type was not taken into consideration for this particular exercise.

3.3.3 Weather Data

Historical meteorological weather data was collected for Pietermaritzburg and was summarised using jupyter notebook as shown in Table 3.3. This data was sourced from the South African Weather Service, and it included minimum and maximum temperatures, as well as precipitation for the past ten years. The temperature is measured in degree Celsius, and the precipitation is the amount of rain daily. The rainfall used for the purpose of the simulations is aggregated annually and is measured in millimetres. The readings shown below are taken daily. The first column shown the record number, zero being the first record.

Table 3.3 Climatic data collected for Pietermaritzburg

Date Time	T(max)	T(min)	P(mm)
2000/01/01	29,33	17,85	32,03
2000/01/02	21,20	15,21	0,00
2000/01/03	22,40	15,78	0,00
2000/01/04	25,08	16,31	0,00
2000/01/05	27,28	17,69	0,00
.....

3.4 Data generation

Data collection for the purpose of the given project required expert knowledge of the required features. Since the information collected was from different sources, and needed to be combined, a data file was generated for the purpose of simulating the ML. A method called feature construction or augmentation was used for this exercise.

Feature construction involves a process of transforming a set of input feature into new features that can be used for prediction purposes [18]. In image classification, the type of augmentation may include a method called random erasing, in which parts of the image are being randomly erased to improved image classification [52]. Data for crop selection was collected from the Department of Agriculture, Land reform and rural development website [53], as already mentioned. The file for simulation was created using the Excel spreadsheet random generator function. The random generator in Excel takes the upper and the lower limit to generate the numbers in between. A total of 150 random numbers were generated per feature to create 150 records per crop.

The climatic features selected are minimum and maximum temperature, minimum and maximum rainfall, and the edaphic features selected are minimum and maximum soil pH and humidity required for production of the selected crops. On the other hand, the averages of the temperature, pH level and rainfall are used. The value of the rainfall is the amount of rainfall distributed over a year while the selected crops are those that are well suited for production in KwaZulu-Natal.

According to the planting calendar, there can be many crops which are suitable to be planted in a given month. For example, in January one can plant Swiss chard, potatoes, lettuce, carrots, cauliflower, Brussel sprouts, broccoli and beans. As a result, the model will need to include more unique features in order to predict the best suited crop for the given meteorological and edaphic circumstances. The labels of the selected features are Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, soil pH level, the amount of rainfall, the month in which the crop may be planted according to the planting calendar and the crop name suited for particular conditions.

3.5 Data analysis

In this section the descriptive statistics of the quantitative data is presented and explained according to the findings of the research as collected by the mentioned questionnaire. This is to understand the case study for the selected area better, and to look closer at the farming community under study.

3.5.1 Demographics of the respondents

Amongst the group of respondents, 60.8% were female, 27.8 were male, and 11.4 of the respondents did not specify their gender as shown in Figure 3.6. This means that this farming community is dominated by female farmers than the male farmers. As shown in Figure 3.7, 16.5% of the respondents were between the ages of 18-35, which is the same percentage as the over 66 group of respondents, and the groups 46-55 and 56-65 groups also had the same percentage of 20.3%. The group 36-45 had the highest percentage of 26.6%. This means that even though the younger generation is beginning to take interest in the farming business, this farming community is dominated by a generation of ages over 45 years constituting over 60% of the group.

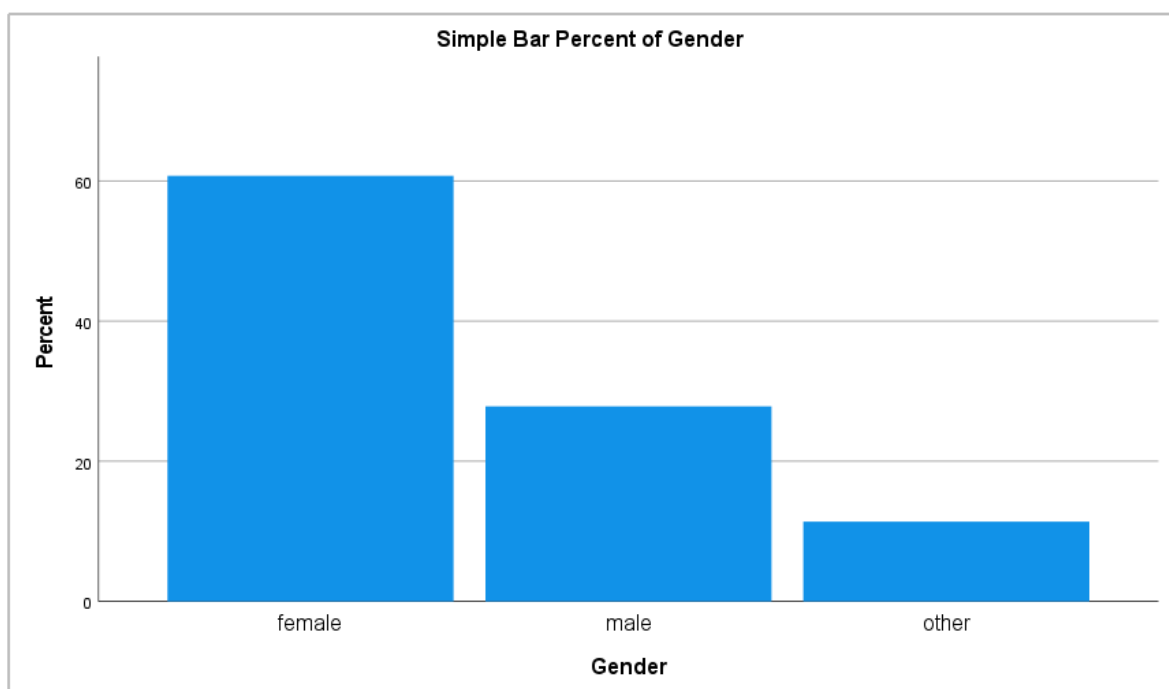


Figure 3. 6 Bar chart of the Gender demographics

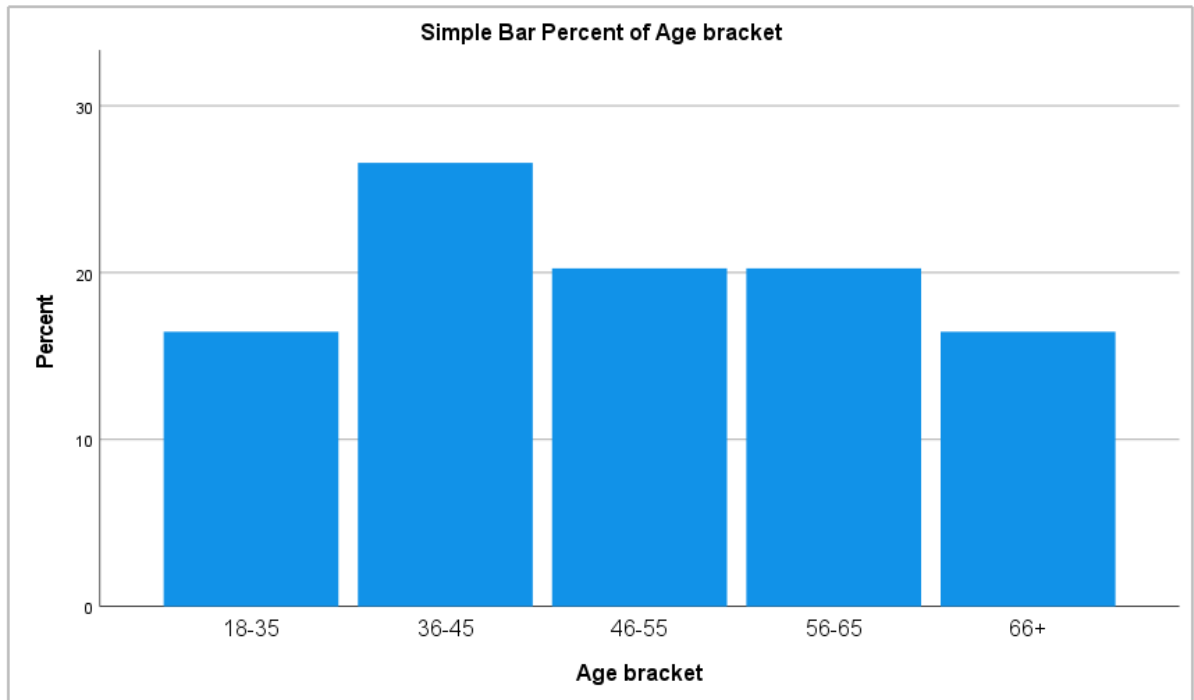


Figure 3. 7 Age groups of the farmers

The largest percentage of 36.7% of the respondents have a secondary school level of education as shown in Figure 3.8, while 15.2% have primary level and 20.3% have no education. The middle percentage of 27.8% has post-secondary education level. This is probably because it has become easier to access schools in the rural areas because almost 80% of the respondents have some level of education, and the secondary and post-secondary education together constitutes more than 60% of the group.

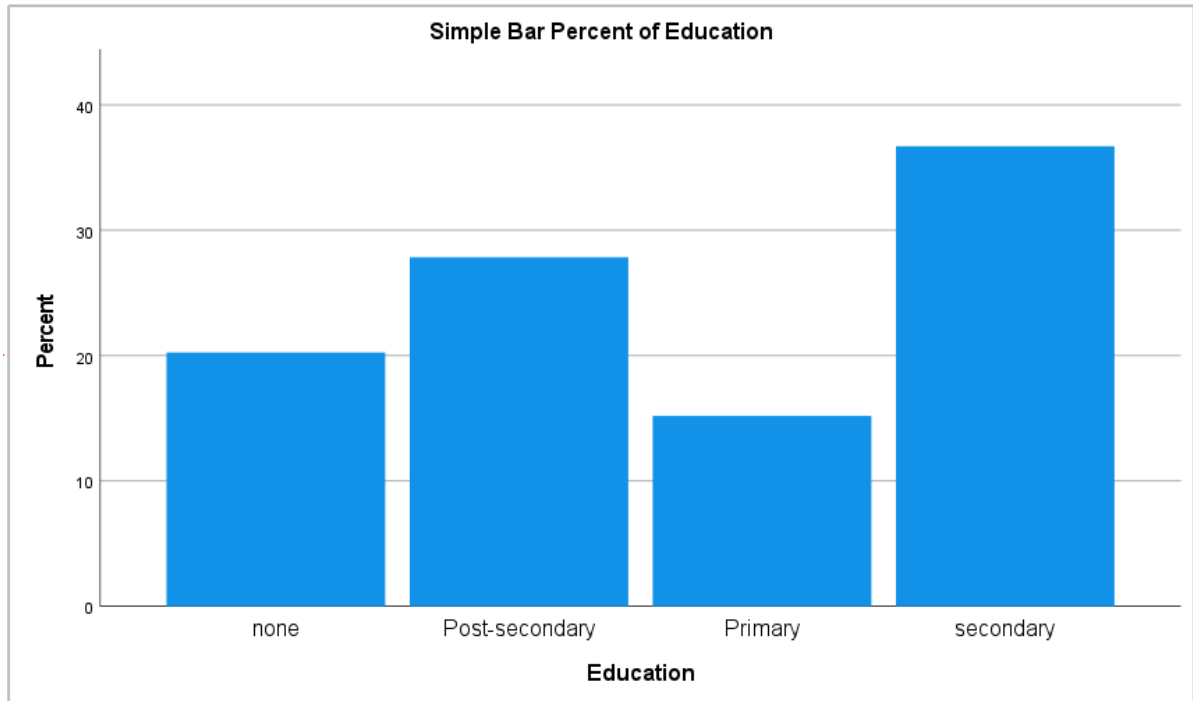


Figure 3.8 Level of Education

When it gets to the number of years of experience in farming, 75.9% of the respondents have more than 20 years of experience as shown in Figure 3.9. This percentage is a combination of all age groups of respondents. This means that there is no limit to the age of starting to be a farmer, and that some of the farmers were probably in the farming business before they even started school.

This means that farming is the core of this community. About 20.3% of the respondents have between 10 and 20 years of experience, leaving a percentage below 5% as having less than 10 years of experience.

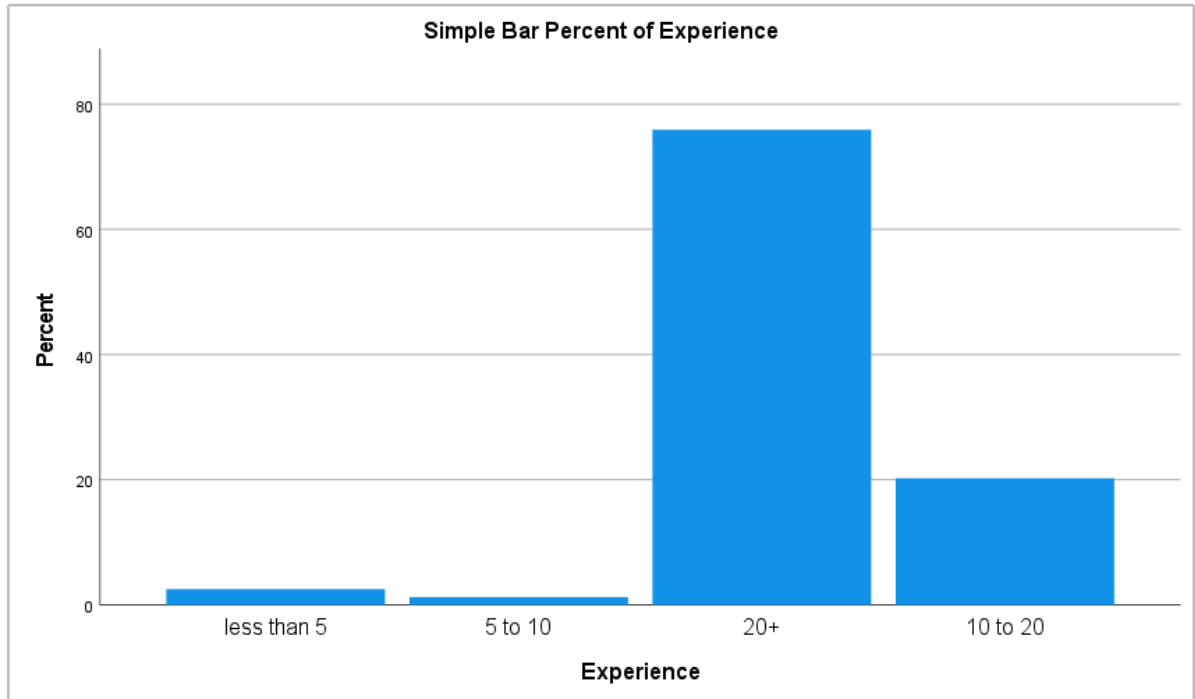


Figure 3. 9 Level of experience in farming

3.5.2 Testing the questionnaire reliability

To test the reliability and consistency of the questionnaire, a section of the questionnaire was used to calculate the Cronbach's Alpha for the four questions that were asked about the subject. The following are the questions selected:

Q1. Using indigenous knowledge, it is simple and reliable to foresee the agricultural season (season with fine weather and good rains) before I cultivate

Q2. Using indigenous knowledge, I can foresee bad season (season with no rains and undesired temperature) before I cultivate

Q3. With Indigenous knowledge in place, there is no need for any form of technology that can help to foresee agricultural season

Q4. Using the indigenous knowledge, it is easy and quick to monitor the development of my crops

Q5. Using indigenous knowledge (observations), I can always spot all factors that may impact my crops such as crop insects and weeds at an early age

The following are responses according to the sample data collected and are represented using bar charts. The y-axis represents the number of responses, while the x-axis shows the respective latent variables which were scaled from 1 to 5, 1 being strongly agree and 5 being strongly disagree.

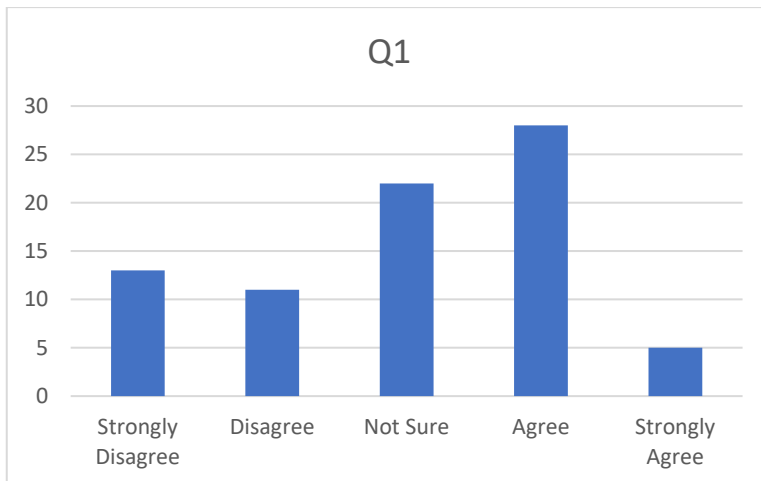


Figure 3.10 Responses for question 1

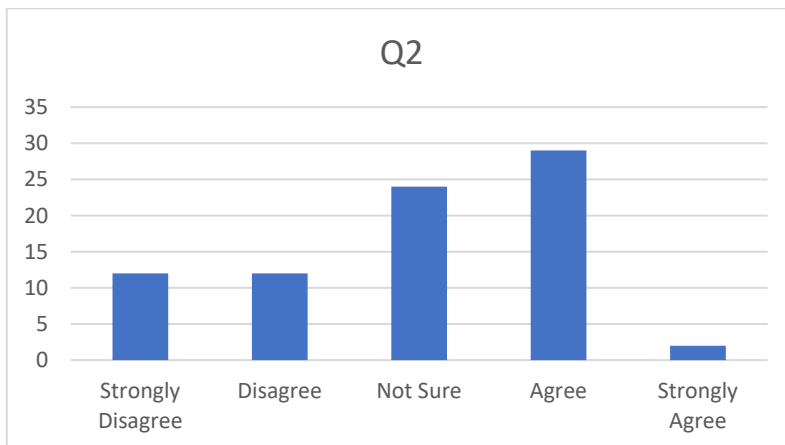


Figure 3.11 Responses for question 2

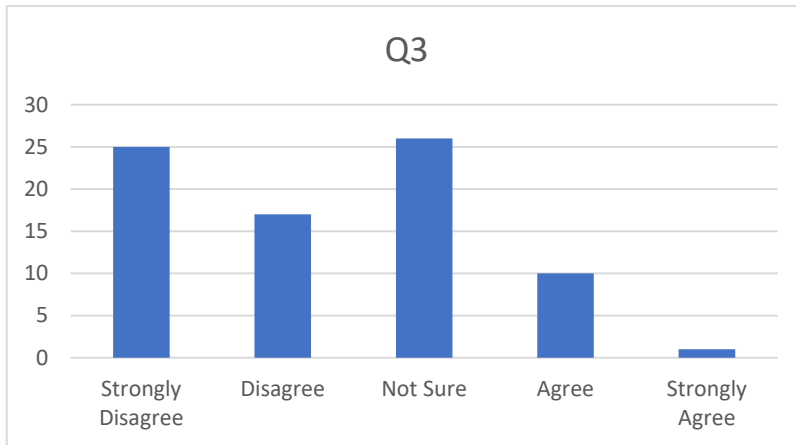


Figure 3.12 Responses for question 3

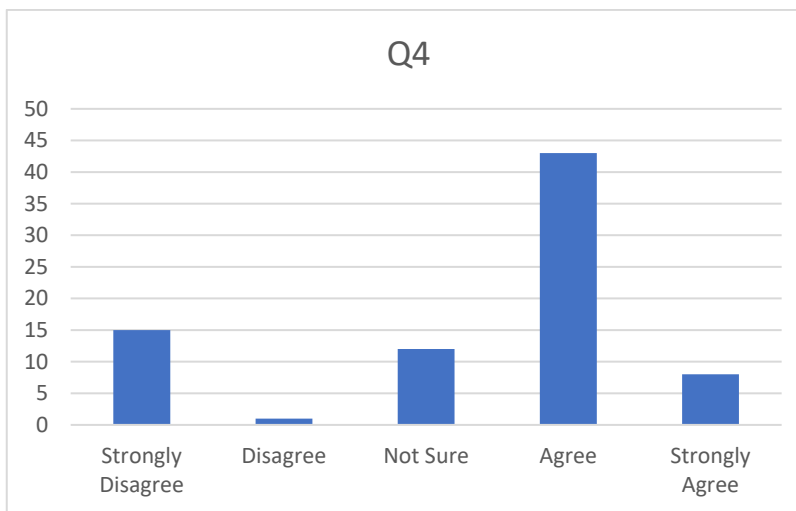


Figure 3.13 Responses for question 4

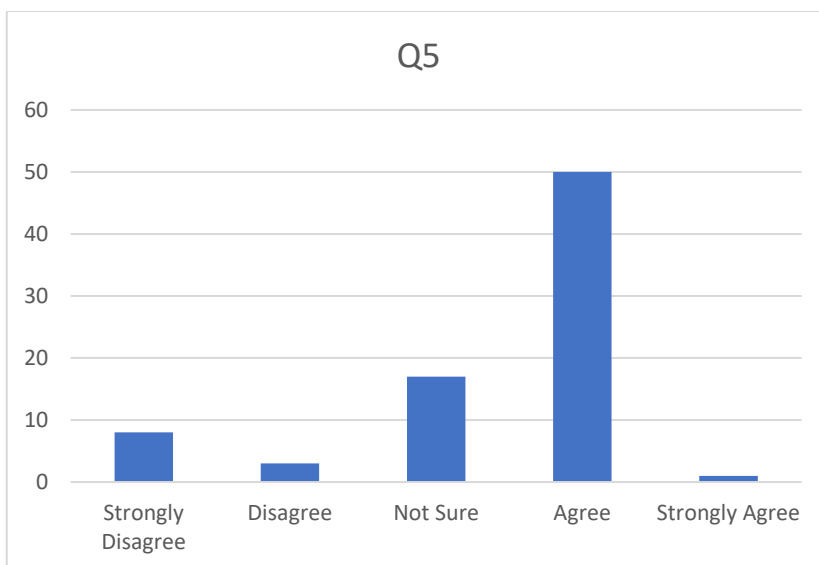


Figure 3.14 Responses for question 5

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Q1	79	1	5	3.01	1.193
Q2	79	1	5	2.92	1.130
Q3	79	1	5	2.33	1.083
Q4	79	1	5	3.35	1.271
Q5	79	1	5	3.42	.982
Valid N (listwise)	79				

Figure 3. 15 Descriptive statistics for the dataset

Case Processing Summary			
		N	%
Cases	Valid	79	100.0
	Excluded ^a	0	.0
	Total	79	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics	
Cronbach's Alpha	N of Items
.972	5

Figure 3. 16 Reliability results for the dataset

The Cronbach alpha value of 0.97 was found. The value that is higher than the value of 0.90 is considered excellent and is an indication of good internal consistency and therefore reliability. This could also be an indication of that some of the variables are too identical and thus the same thing reoccurs, but it is not the case in this instance. Even when the number of items were reduced, the reliability coefficient remained above 0.90. When one looks at the questions individually, it becomes clear that the small-scale farmers do believe in the help of indigenous knowledge hence a higher percentage agree with the positive idea that indigenous knowledge is critical to their daily operation.

3.6 Evaluation procedures

Machine learning process was performed on the climate and edaphic data taking into consideration the type of problem to be solved, the size of the dataset and how the size would be split between training and testing, and lastly the number of features. The selected algorithms were employed and evaluated according to machine learning classification metrics for evaluation. Metrics used for classification evaluation include accuracy score, precision, recall, F1-score, ROC, and AUC. The metrics selected in this case to compare and contrast the algorithms are precision, recall, F1-score and the accuracy score. The best possible algorithm was then selected. The process is described in more detail in Chapter 4.

3.7 Conclusion

In this chapter, the systematic approach of this research was presented. The elements of a research methodology were demonstrated. The first element covered is measurement objectives, that is the reason for the objectives and the envisioned outcomes were explained. The other element covered is the data collection process. It was explained how the data was collected, methods used for the data collection, and the type of data that was collected. This chapter also explained how the data was used to achieve the results that will be explained in later chapters. The description of the sample area and the sample population was also provided in this chapter. The data analysis methods for the collected data were also described in this chapter. In conclusion, the framework for the intelligent agro-climate decision support system was explained. The next chapter will explain the ML evaluation process and the findings of the process followed.

Chapter 4: Modelling of the Intelligent Agro-climate Decision Support system

4.1 Introduction

Crop agriculture by small-scale farmers depends on factors such as the economy, the demand, access to seeds, farm implements, and access to markets. There are two types of decisions that farmers have to make in the crop agricultural maintenance. Long-term goals to develop and strengthen the firm, such as marketing, are referred to as strategic decisions. Planting timing, crop suitability, and crop variety selection are all tactical decisions. The small-scale farmer's tactical decisions will be used as a bedrock for the modelling of the Agro-climate Decision Support system. The aim of this chapter is to outline simulations and activities to the modelling of the Intelligent Agro-climate Decision Support system (DSS).

This research will concentrate on crop suitability in relation to climatic and edaphic conditions. The climatic and edaphic conditions of the selected area in KwaZulu-Natal have been used in these data simulations. The study utilises Machine Learning (ML) to determine the suitable crop to be planted at a particular time according to the specified climatic and edaphic condition. A sequential ML process will be followed, and the results analysed and presented. At the second phase, Indigenous knowledge (IK) will be introduced to collaborate and compare the results. The chapter focuses on ML which provides the intelligent part of the DSS. Jupyter Notebook from the Anaconda3 package was used for this ML exercise in a quest to design the perfect model for crop selection at planting stage for small-scale agriculture.

4.2 Machine learning

Machine learning in agriculture has been widely used in crop management mainly for prediction of crop yield [45]. Machine learning is a scientific field that allows machines to function without being programmed [54]. In a ML process, machines go through a process of learning or training in task performance before they can be used to make predictions. The performance of the ML model is measured by making use of a metric.

The learning process is dependent on different algorithms that are employed. The process of learning can either be supervised or unsupervised, and the learning model is selected according to the task at hand. In supervised learning data is presented with examples that define specific rules. There is distinction between data used for training the model and the data used to test the model. In terms of unsupervised learning, there is no input that defines the rules, but the model is used to define or discover its own patterns. Some of the ML models that are used in agriculture include regression models, clustering, Bayesian model, instance based, artificial neural networks, ensemble, support vector and decision trees. A ML project follows particular steps as already mentioned in the previous chapter. In this chapter, the steps followed for the ML process are discussed in more detail.

4.2.1 Data collection

Data collection for the purpose of the given project requires expert knowledge of the required features. A brute-force method of data collection was used for this exercise. Data for crop selection was collected from the Department of Agriculture, Land reform and rural development website [53]. The information is documented in the different brochures that are created with production guidelines. Each crop requirements are documented separately with information that include description, climate requirements, soil condition, rainfall requirements, nitrogen, phosphorous and potassium requirements, preparation, and harvesting. A technique called feature construction was used to create a randomised data file for simulations from the information gathered using random number generator on excel. The random generator in excel takes the upper and the lower limit to generate the numbers in between. With each crop, 150 random numbers were generated per feature. Table 3.1 is a sample of the selected crops and the features that are documented. The information was gathered from a web page [53].

The climatic and edaphic feature selection was already discussed in the previous chapter. It is imperative to note that for the purpose of the modelling task, the temperature was aggregated, and the value of the rainfall used is the amount of rainfall distributed over a year while the selected crops are those that are well suited for production in KwaZulu-Natal. The selection was informed by the KwaZulu-Natal planting calendar. The calendar

is published by seeds for Africa for every province in South Africa. The mentioned planting calendar for the area was acquired from the website of DALRRD [51]. The selected crops are the ones that are typically planted by the Pietermaritzburg small-scale farmers. The soil type was not taken into consideration for this particular exercise. Figure 4-2 provides the KwaZulu-Natal planting calendar.

Table 4. 1 Planting chart for KwaZulu-Natal

Vegetable Planting Chart Kwazulu Natal												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Beans (Broad)				■	■							
Beans (Bush)	■							■	■	■	■	■
Beans (Climbing)	■							■	■	■	■	■
Beetroot	■	■	■	■				■	■	■	■	■
Broccoli	■	■										■
Brussel Sprouts	■	■										■
Cabbage								■	■	■		
Carrot	■	■	■	■				■	■	■	■	■
Cauliflower	■	■	■									
Celery		■	■	■								
Corn								■	■			
Cucumber	■							■	■	■	■	■
Eggfruit (Brinjal)								■	■	■	■	■
Lettuce	■	■	■					■	■	■	■	■
Melons								■	■	■	■	■
Onion		■	■									
Parsnip								■	■	■		
Peas				■	■	■						
Peppers									■	■	■	
Potatoes	■	■							■	■	■	■
Pumpkin								■	■	■	■	■
Radish							■	■	■	■		
Swiss Chard / Spinach	■	■	■	■				■	■	■	■	■
Tomato								■	■	■	■	■
Watermelon								■	■	■		

According to the planting calendar, there can be many crops which are suitable to be planted in a given month. For example, in January one can plant Swiss chard, potatoes, lettuce, carrots, cauliflower, Brussel sprouts, broccoli and beans. As a result, the model will need to include more unique features in order to predict the best suited crop for the given meteorological and edaphic circumstances. The edaphic features that were taken into consideration were the Nitrogen (N), Phosphorus (P), Potassium (K), and soil pH level. The climatic features taken into consideration were rainfall, and temperature. The features are summarised with labels and the first few records are shown in Table 4.2. The

headings show the amount of rainfall, the month in which the crop may be planted according to the planting calendar and the crop name suited for particular conditions.

4.2.2 Data analysis

As mentioned previously in this chapter, the data that was used was created using brute force. The Python API was used to load the generated excel file to view and manipulate the data. To examine the data, the first few rows were reviewed with headings as seen in Table 4.2. The table shows the first five records from the excel file in a table form with the respective headings. It is imperative to know how many rows and columns are available to work with. There should not be too many features because some algorithms may suffer poor performance, and too many rows may result in training taking too long.

Table 4.2 Contents of the Crop simulation file

	N	P	K	Temperature	Humidity	pH	Rainfall	Month	Crop
0	71	54	16	22.613600	63.690706	5.749914	87.759539	8	Maize
1	61	44	17	26.100184	71.574769	6.931757	102.266244	9	Maize
2	80	43	16	23.558821	71.593514	6.657965	66.719955	8	Maize
3	73	58	21	19.972160	57.682729	6.596061	60.651715	8	Maize
4	61	38	20	18.478913	62.695039	5.970458	65.438354	8	Maize

The statistical description of the given file is represented in Table 4.3. It is a summary of the file contents. The standard deviations of all the features are exceptionally large, with the exception of the soil pH value. This means that the soil pH values of the different records do not deviate much from the average value of the soil pH values in the dataset. The minimum of all the values is greater than zero, meaning that there were no missing values in this dataset. The value of the range of the features is extremely high with the exception of the soil pH value. This means that the other features are widely spread out and less consistent.

Table 4.3 Statistical description of the crop simulation file

	N	P	K	Temperature	Humidity	pH	Rainfall	Month
count	750.000000	750.000000	750.000000	750.000000	750.000000	750.000000	750.000000	750.000000
mean	87.169333	88.845333	120.633333	23.034610	64.842701	6.077383	273.206183	8.517333
std	52.544666	65.513745	58.763427	3.934903	25.021230	0.576462	265.399567	3.083890
min	10.000000	11.000000	15.000000	15.330426	18.092240	5.004221	21.087703	1.000000
25%	45.000000	43.000000	103.000000	20.652426	54.851929	5.638354	65.438354	8.000000
50%	76.000000	68.000000	129.000000	22.476291	67.225063	6.138617	109.062734	9.000000
75%	115.000000	123.750000	162.000000	24.132438	87.169779	6.575932	511.685402	10.000000
max	200.000000	250.000000	240.000000	37.833084	94.993309	6.997326	906.177928	12.000000

Data types may also be reviewed for the features used. Machine learning makes use of numeric values and if the values are not numeric, there may be a need to code them beforehand. If the data has fields such as date, there would also be a need to extract the required field from it. The descriptive statistics according to Table 4.3 show the total number of records, the average per feature, the mean, and standard deviation. There are seven hundred and fifty records and eight features, with the last column showing the category. The data is organised according to labels. This means that supervised machine learning can be employed for this dataset.

Supervised ML can be done using classification or clustering. As already mentioned in the previous chapters, ML solutions depend on the kind of problem that is being addressed. We can conclude that the problem addressed in this case is a classification problem. The goal is to categorise the given data into categories of crops available. There is therefore significance in determining how balanced the class/category values are. In a case where there is no balance, special preparation would be required to avoid bias towards the majority class. This means that if one class had more records than the other, the results could be biased in favour of the class with more records because it would be predicting that class more times than the rest.

The heat map in Figure 4.1 confirms there are no records with missing values. The Count plot shown in Figure 4.2 confirms number of records per crop name as 150. This means that the dataset is balanced, meaning there is not one class with more records than the

rest, and there are no gaps in the data. The dataset shows no missing values, meaning there are no records with fields that have any missing values. If there were missing values or the number of records were not the same, further processing would be necessary to eliminate the imbalance. Figure 4.2 show that the class values are all balanced, meaning they have the same number of records. It confirms the narrative in Figure 4.1.

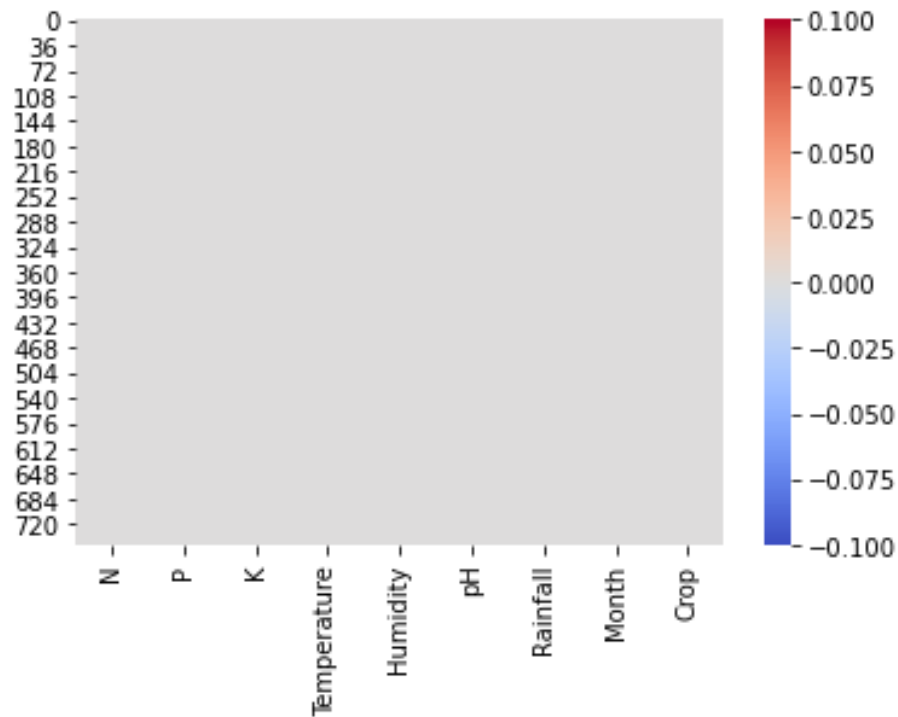


Figure 4.1 Heat map of the contents of the crop simulation file

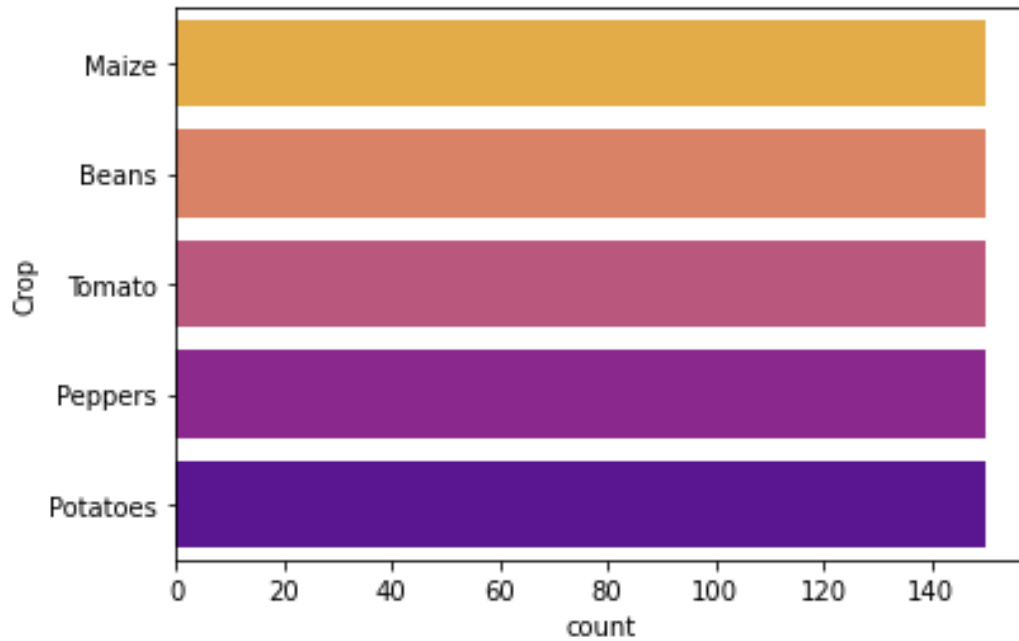


Figure 4.2 Count Plot of the data distribution

The comparison between the temperature and pH distribution in Figure 4.3 shows the temperature to have a more symmetrical shape. This means it is a Gaussian distribution for the temperature data. It is also shown in Figure 4.5 that there is a strong negative correlation between the two features. This suggests a negative relationship between the two.

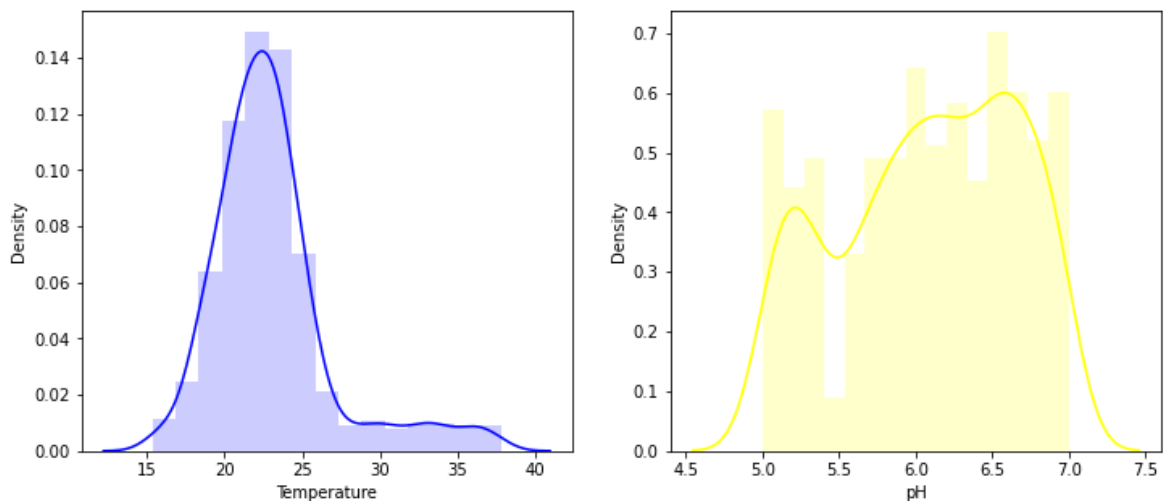


Figure 4.3 Comparison between Temperature and PH

The comparison between the temperature and humidity shown in Figure 4.4 on the other hand seems to produce a different result. This comparison shows a lot of outliers with the humidity. In Figure 4.9 there is a strong positive correlation between the temperature and humidity. This suggests a relationship between the two features.

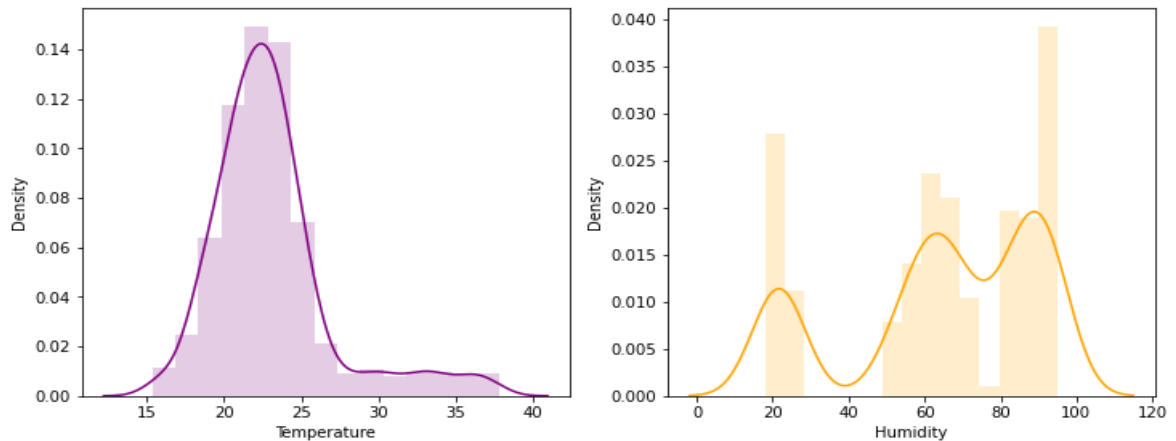


Figure 4.4 Comparison between Temperature and humidity

Heat map of Correlation of the features is shown in Figure 4.5. The straight diagonal line shows a perfect negative correlation of the features. Figure 4.6 and Figure 4.7 show the distribution of each figure in a histogram and density plot, respectively. Correlation between attributes as shown in Figure 4.5 is important because some algorithms may suffer poor performance if there is high correlation between the features. If the problem that is being solved was a regression problem, the perfect correlation would pose a problem for regression model and would produce misleading results.

Visualization of correlation between features

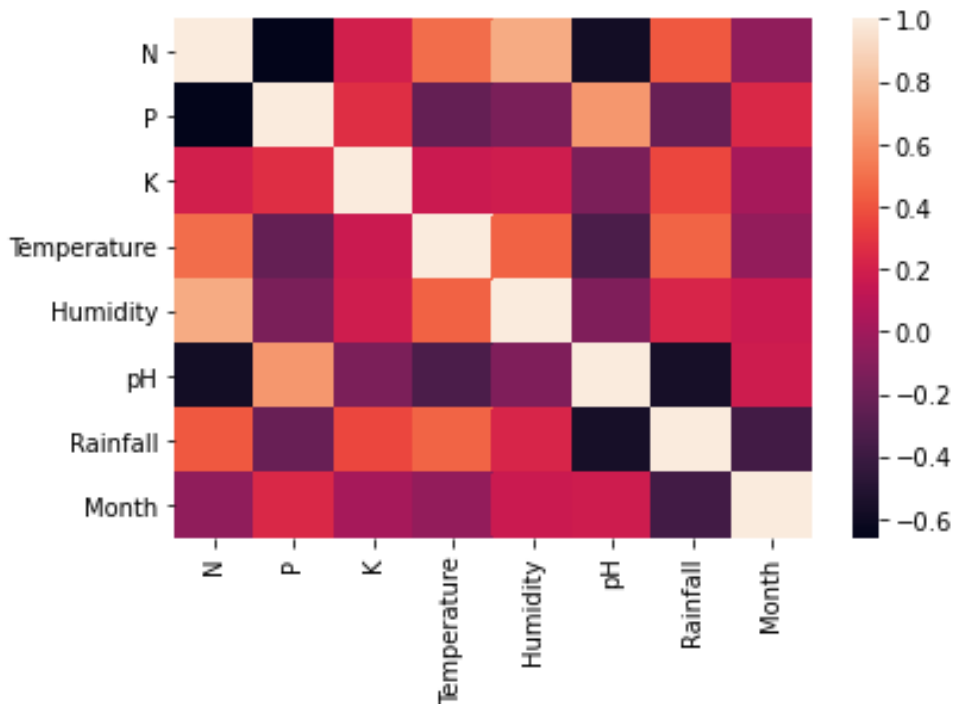


Figure 4.5 Heat map of Correlation between features

Histogram plots and Density plots are shown in Figure 4.6 and 4.7 respectively to determine the distribution of each feature. Many machine learning techniques assume a Gaussian variate distribution. From the figures it can be determined that none of the features exhibited a Gaussian distribution except for the temperature feature, leading to a decision to normalise or standardise the data before working with it.

Figure 4.9 is a box and whiskers plot for the distribution as well, with the difference that this plot draws a line through the median and a box around the 25th and 75th percentile of the distribution. The whiskers show how the data is spread and the black dots outside the whiskers show outliers, as seen with the Month, Temperature and Phosphorous Features. For the month feature it is because amongst the crops that are involved, only two crops may still be planted in the months January and February.

Figure 4.8 show a scatter plot of the distribution for humidity and rainfall. A scatter plot is a visualisation of the relationship between two values. In the mentioned figure, the

relationship between the humidity and rainfall is shown for the different classes. The different classes are depicted using the different colours. The same figure (Figure 4.8) shows how the different classes have completely different values for the mentioned features.

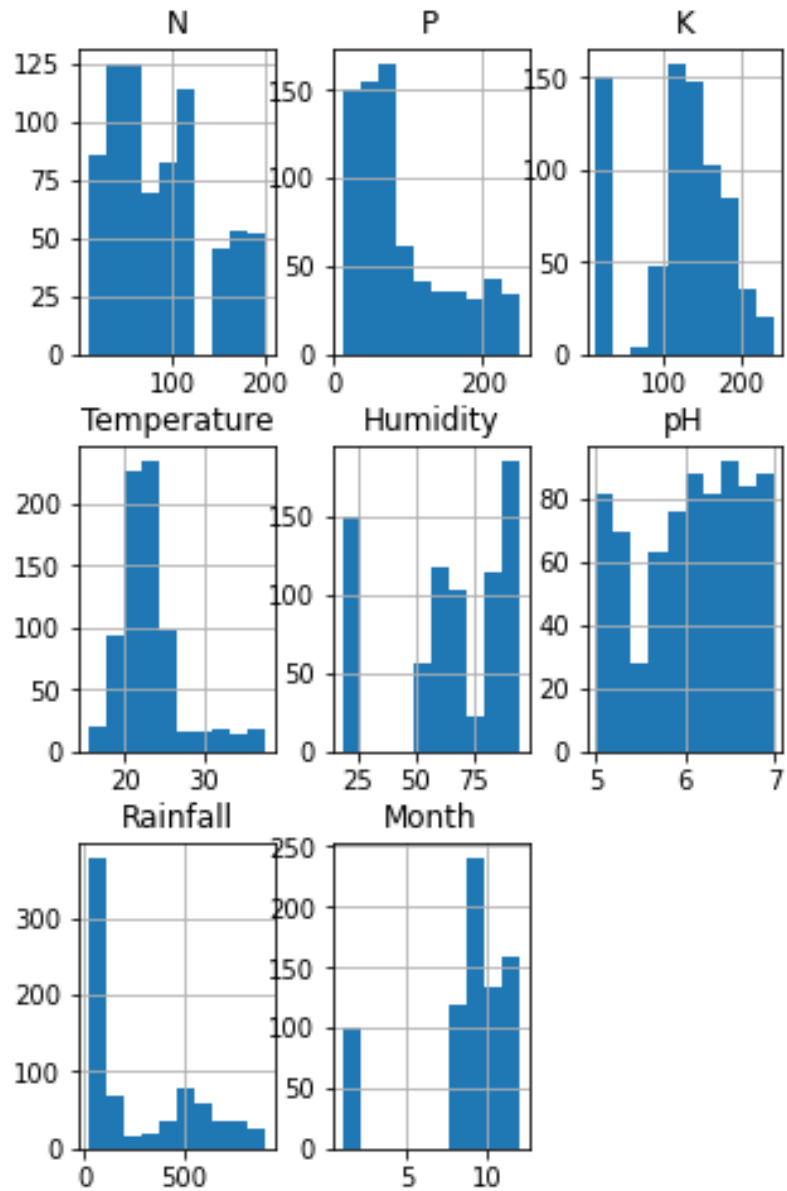


Figure 4.6 Histogram of each of the features

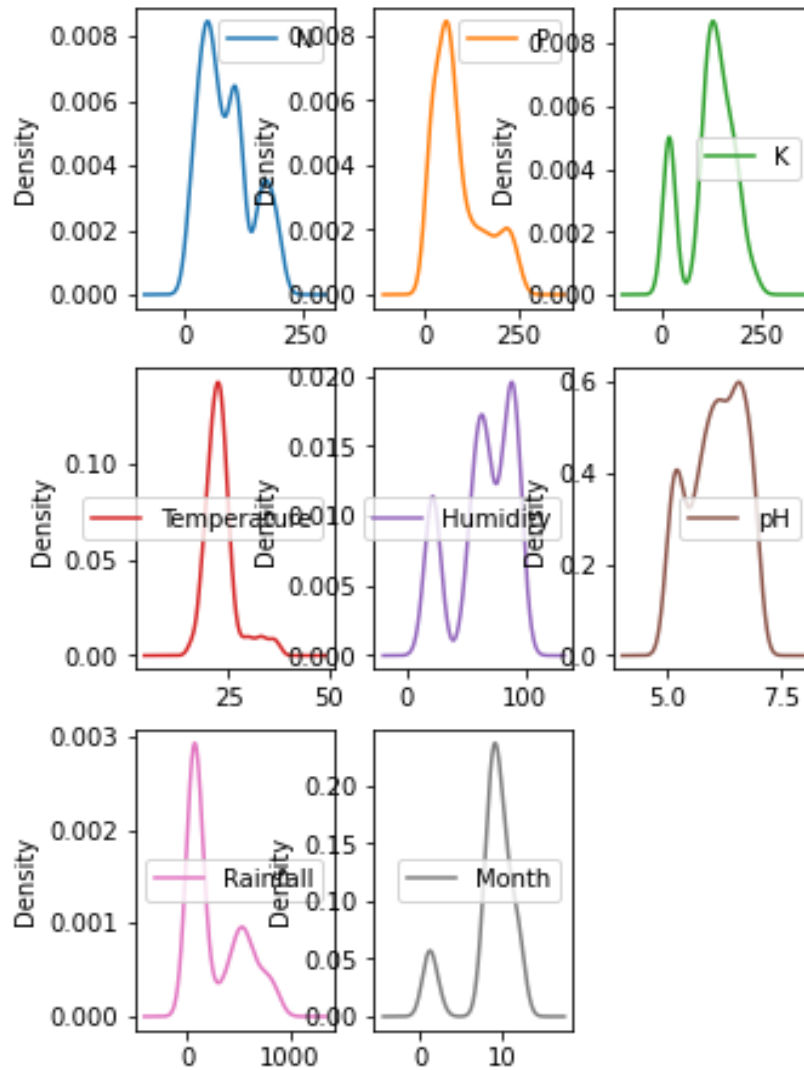


Figure 4.7 Density plot of each feature (the important part is the curve/shape of the line)

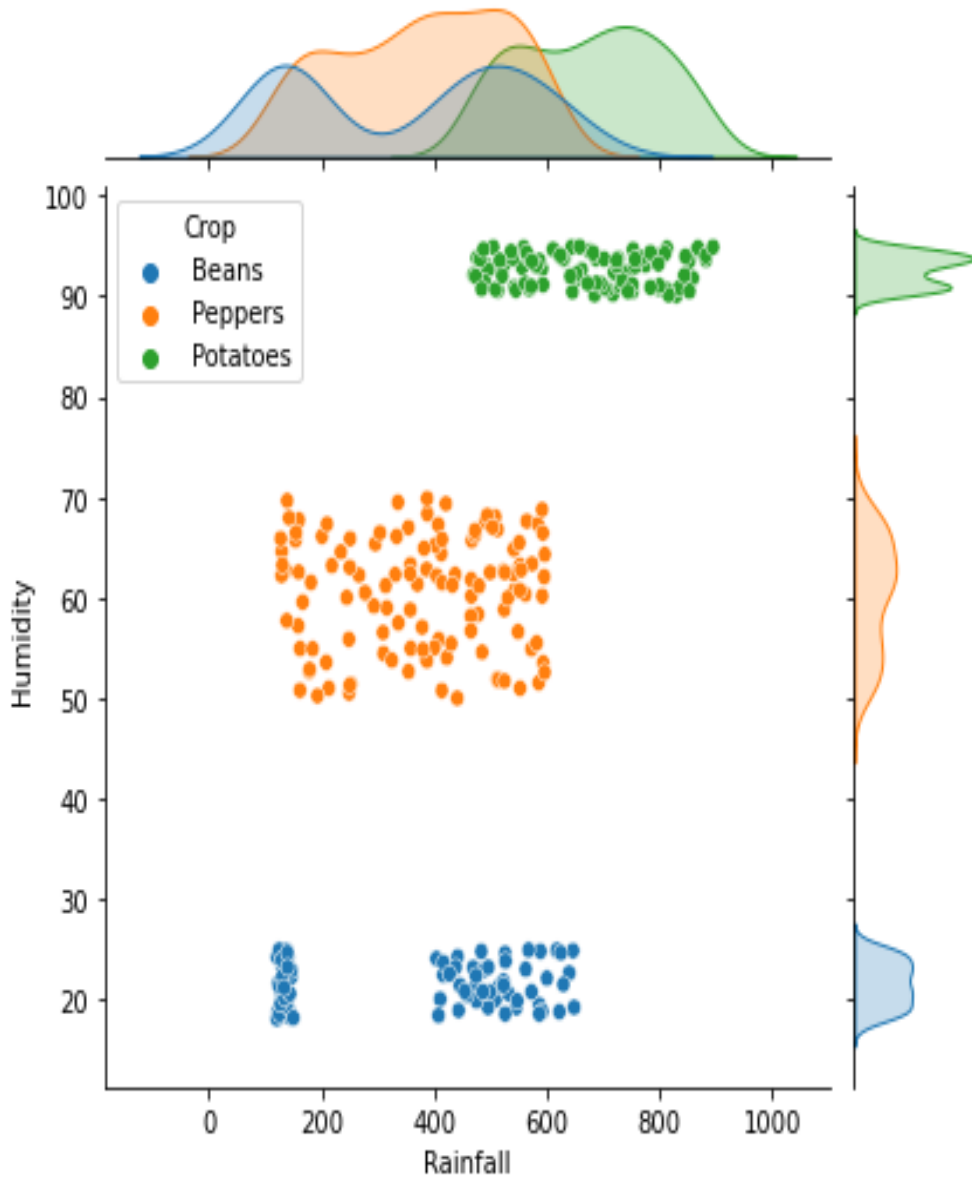


Figure 4.8 Joint plot of the rainfall against humidity shown for 3 crops

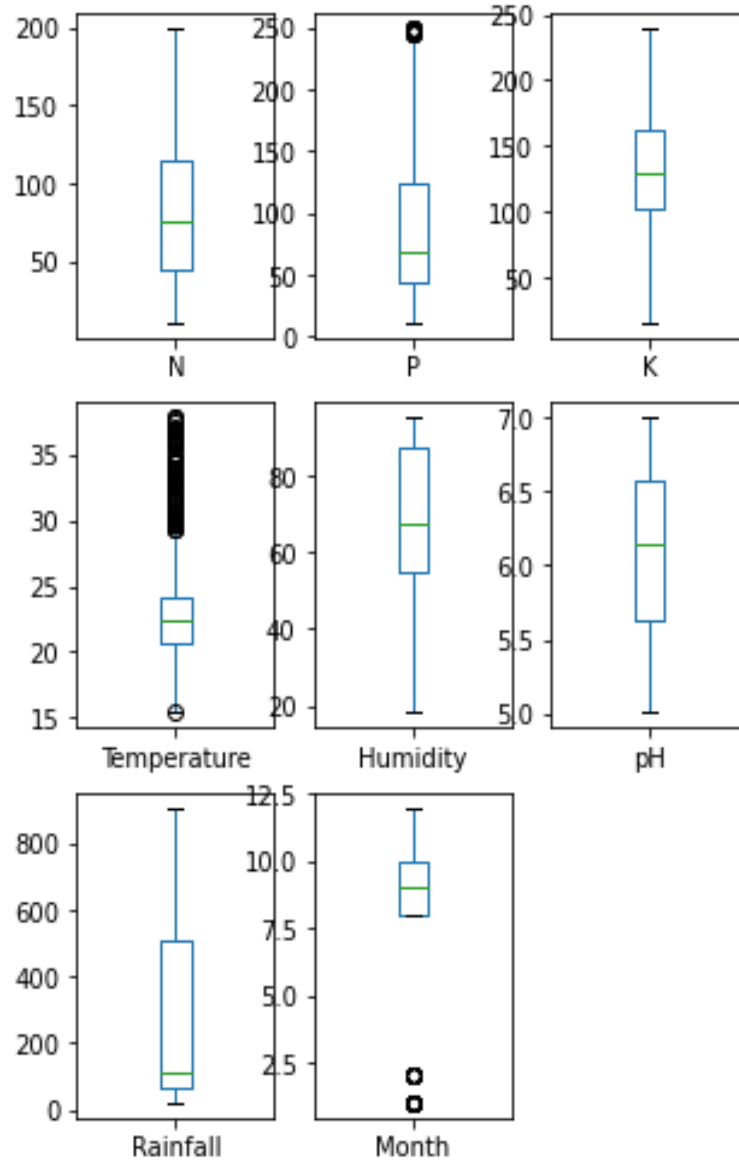


Figure 4.9 Box and whiskers plot for each feature

The data shows that there is a strong correlation between the climatic features and other climatic features, but so much correlation to the edaphic factors. For example, there is a strong correlation between rainfall and pH of the soil, and pH has high correlation to Nitrogen. None of the other edaphic features have high correlation to climatic features.

4.2.3 Data pre-processing

Feature scaling, often known as standardisation, is a data pre-processing process applied to independent variables or data characteristics. It essentially aids in the normalisation of data within a specific range. Machine learning methods that use Euclidean Distance measure to determine distances between data require feature scaling. If scaling is not done, when computing distances, the feature with the higher value range takes precedence. The assumption behind feature scaling is that your data follows a Gaussian (bell curve) distribution. The following methods are normally used for feature scaling [55]:

a) Rescaling (min-max normalisation)

The simplest method of feature scaling. It is also known as min-max scaling or min-max normalisation. It entails rescaling the range of features to scale the range in [0, 1] or [1, 1]. The goal range is determined by the data type. The following is the general formula for a min-max of [0, 1]:

a) Rescaling (min-max normalisation)

$$x^1 = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3)$$

The following formula in equation 4 applies:

b) Mean normalisation

$$x^1 = \frac{x - \text{average}(x)}{\max(x) - \min(x)} \quad (4)$$

Feature standardisation gives each feature in the data a zero-mean value. The value of the standardised variable is calculated as below.

c) Standardisation (Z-score Normalisation)

$$x^1 = \frac{x - \bar{x}}{\sigma} \quad (5)$$

Another option that is widely used in machine-learning is to scale the components of a feature vector such that the complete vector has length one. This usually means dividing each component by the [Euclidean length](#) of the vector as in the formula below.

d) Scaling to unit length

$$x^1 = \frac{x}{\|x\|} \quad (6)$$

As already said, because many machine learning algorithms make assumptions about the data, there is a need to transform the data so that it can be usable. Because Gaussian distribution is assumed, standardisation was performed on the data as shown in Figure 4.15. As already mentioned in point number c), the formula shown in equation 5 applies. The MinMaxScaler from sklearn library was used to transform the data for processing to avoid getting misleading results using the train_test_split function.

4.2.4 Choosing an algorithm

The algorithm chosen should be the most appropriate for the type of problem to be solved. The features of the dataset have precisely defined labels based on the data that has been collected. The task at hand is to categorise instances of the given data into one of many crops. Supervised machine learning algorithms can be used to complete the given problem. Examples of supervised learning algorithms include but are not limited to Linear Classifiers, Logistic Regression, Naïve Bayes Classifier, Perceptron, Support Vector Machine, K-Means Clustering, Boosting, Decision Tree, Neural networks, and Bayesian Networks. The goal is to anticipate a suitable crop to plant at a specific time based on the given edaphic parameters. As a result, a logic-based classification method is recommended for this purpose. Logic-based Learning entails the automated creation of

logic-based algorithms from examples and domain knowledge. Decision trees and rule-based classifiers are some of the logic-based algorithms. Classification trees that categorise instances by ordering them according to feature values are known as decision trees.

4.2.5 Training the dataset

The data obtained is appropriately labelled as already mentioned and contains fewer than 100k records, meaning the problem to be solved is a classification. The classification challenge entails determining which category an object falls into. An object is a data item or record in this context, and it is fully represented by a set of values or features. Categorisation can only predict values in a distinct set. In this scenario, the goal is to forecast which crop to plant based on climate and edaphic conditions. There are four categories of classifications to choose from: binary classification, multi-class classification, multi-label classification, and imbalanced classification. Those classification tasks with two class labels are referred to as Binary classification. Those classification tasks with more than two class labels are referred to as multi-class classification. Classification tasks with two or more class labels, where one or more class labels can be anticipated for each case, are referred to as multi-label classification. Imbalanced classification refers to problems in which the number of examples in each class is distributed unequally.

The endeavour is a multi-class classification since the objective is to distinguish which of many categories a single record in our data belongs to. It is therefore imperative that the following algorithms are explored: k-Nearest Neighbours, Decision Trees, Support Vector classifier, Random Forest, and Gradient Boosting. A comparison of all the mentioned classification algorithm follows below. A machine learning map for the process of choosing an algorithm is shown in Figure 4.10.

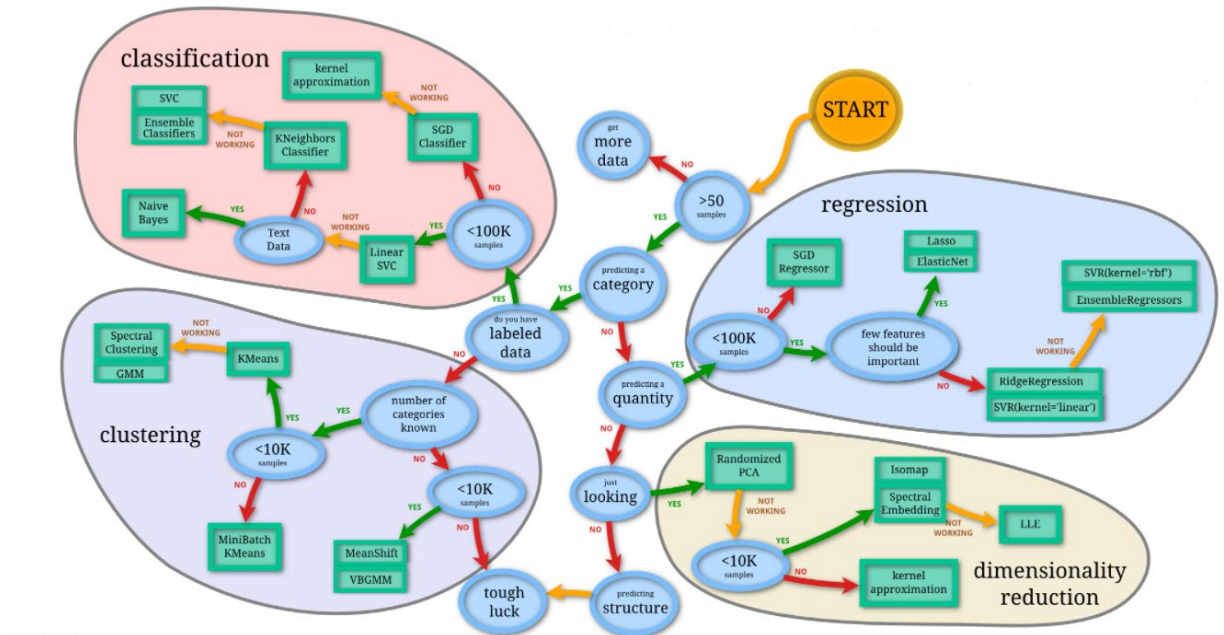


Figure 4.10 Machine learning process for Algorithm selection[56]

Table 4.4 depicts the libraries NumPy, Scikit-learn, Matplotlib, and Pandas that were used in the task. The data that was generated as mentioned was imported into Jupyter Notebook pandas data frame using the “read.excel()” function. The loaded data was split using the train-test-split function from sklearn.

Table 4.4 Machine learning libraries used

Name of the library	Purpose of the library
1. NumPy	For multidimensional array manipulation
2. Pandas	For data manipulation and analysis
3. Scikit-learn	For access of machine learning algorithms and tools
4. Matplotlib, Seaborn	For data plotting
5. Yellowbrick	For visual representation of reports

i. Classification using Decision Trees

Figure 4.12 shows the Decision tree classifier used to train seventy percent of the dataset with a thirty percent used for testing. The sklearn library in Python was used to accomplish the task. When training the model using the decision tree classifier, an accuracy score of 1.0 was observed on training set as shown in Figure 4.12. Accuracy is the measure of how often the classifier classifies correctly overall. This means a 100% accuracy on the training data. On the test data, the decision tree classifier obtained a 99% accuracy, meaning the model is able to predict the results at an accuracy of 99%.

Figure 4.11 shows the feature importance according to the Decision Trees classifier. Feature importance is a measure of how useful an input feature is in predicting the targeted variable feature. Rainfall, Humidity, Potassium and Phosphorous are the features that have been considered more important in the prediction of the target according to this algorithm. Feature importance provides insight as to which features can be eliminated in the process of improving the performance of the algorithm should there be a need to reduce the number of input features. This means that the re-evaluation of the model performance may be based on the feature importance.

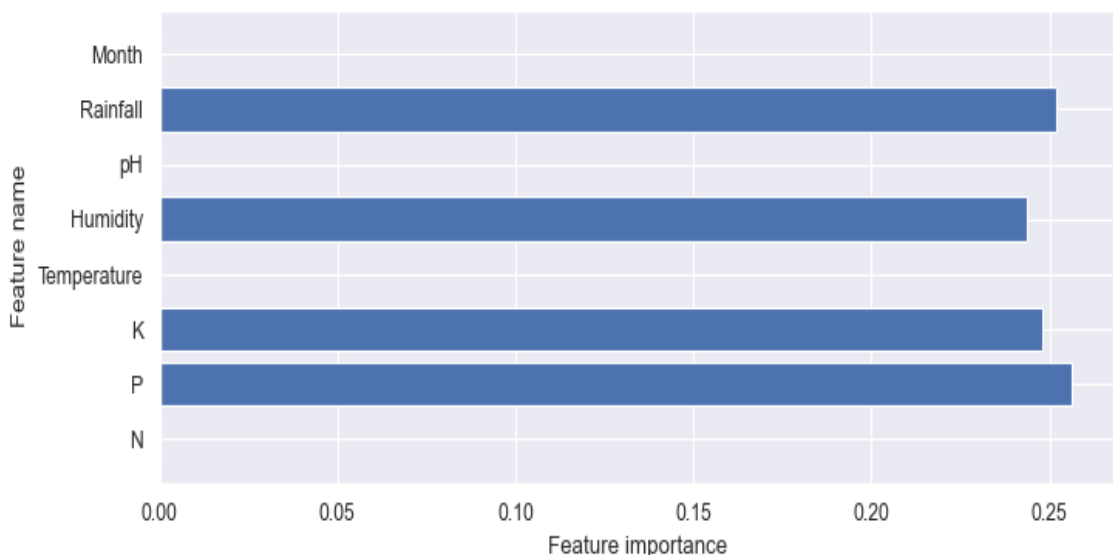


Figure 4.11 Feature importance according to Decision Tree Classification

Figure 4.12 shows the Decision Tree according to the Decision Tree classifier. The decision tree is a schematic representation of the flow of the decision-making process by making use of leaf nodes and branches. The leaf nodes represent a decision outcome, while the branches represent coexistence of features that led to the decision or the decision rules. Decision trees follow a Sum of Product representation, also known as Disjunctive Normal Form. The main challenge a decision tree classifier has is the selection of the root node.

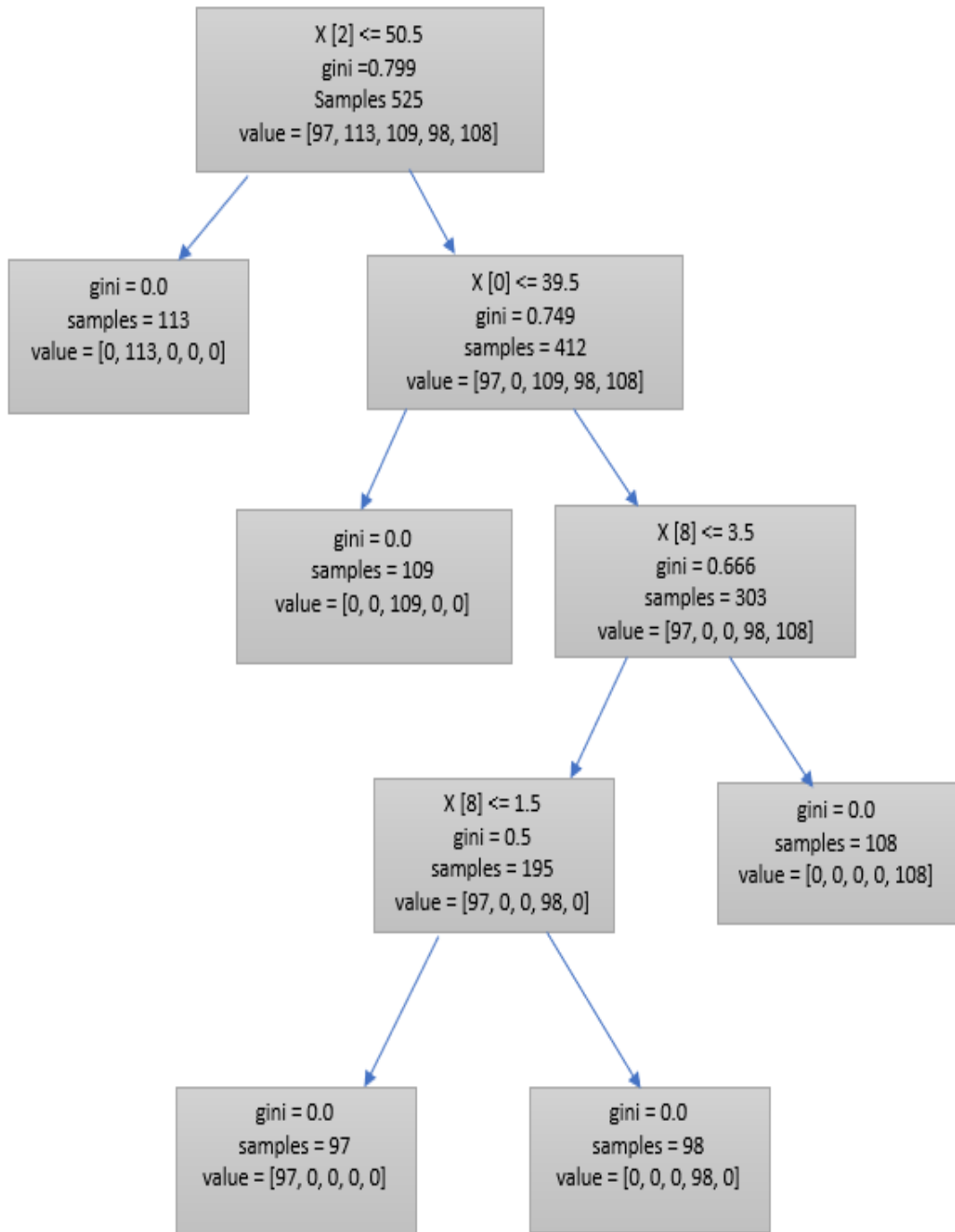


Figure 4.12 Decision tree according to the Decision tree classifier

Figure 4.13 shows a classification report for the Decision tree classifier. There are four ways to check if predictions are right or wrong. The first is True Negative (TN), where the case was negative and predicted negative. True Positive (TP), where the case was

positive and predicted positive. False Negative (FN), where the case was positive but predicted negative. False Positive (FP), where the case was negative but predicted positive. The classification report uses metrics Precision, Recall, f1 and support for representation.



Figure 4.13 Classification report for the Decision tree classifier

Precision is the ability of the classifier to not give a FP, and is measured as the ratio of TP over the sum of TP and FP. The capacity of a classifier to discover all positive cases is known as recall, calculated as the ratio of TP to the sum of TP and FN for each class. Recall is also known as the TP rate or sensitivity. The F1 score is the weighted mean of precision and recall, with 1.0 being best case and 0.0 being worst case. F1 scores are lower than accuracy figures because they factor on precision and recall. The number of actual occurrences of the class in the dataset provided is known as support. Support is used for diagnosis of the evaluation process and does not undergo alteration from model to model.

The report above shows a Precision of 100% for all classes except for Peppers, where only 95.6% of all the positive Pepper instance predictions were actually correct. The Recall for Tomatoes is 0.958, meaning 95.8% of the positive pepper instances were correctly classified. The classes Tomatoes and Peppers have an f1 score of 97.9% and 97.7% respectively, meaning that they were affected by at least one of the factors used, either Precision or Recall.

ii. Classification using K-Nearest Neighbour (KNN) classifier

Figure 4.15 shows results of the training using the KNN classifier. It shows an accuracy score of 100% on the prediction using the test set. This is only when the data has been standardised (scaled). The accuracy of 27% and 30% respectively is shown in the figure when the data is not standardised. This is to demonstrate the importance of preparing the data so that the algorithm can be able to consume it.

Figure 4.14 shows a confusion matrix for the KNN classifier. The confusion matrix is a visual depiction of the actual values against the projected values. The y-axis of the confusion matrix shows the actual values, and the x-axis shows the predicted values. The confusion matrix is also used to measure the performance of a machine learning classification algorithm. The confusion matrix also uses TP, TN, FP, and FN values. The numbers in the diagonal line show how often the actual instance and the predicted instance are the same. The zero values indicate how often the actual instance was predicted incorrectly. This means that the zero values indicate where the errors would be.

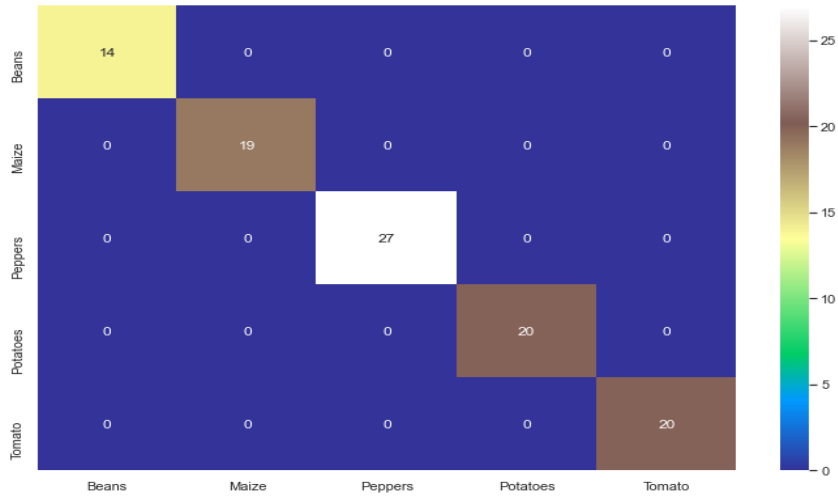


Figure 4.14 Confusion Matrix for KNN classification

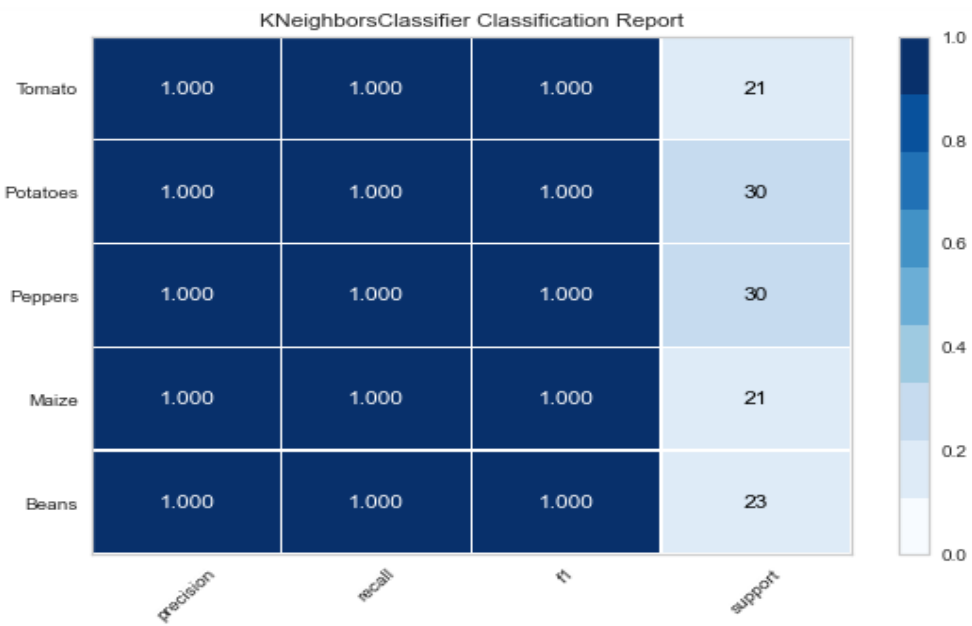


Figure 4.15 Classification report for KNN classification

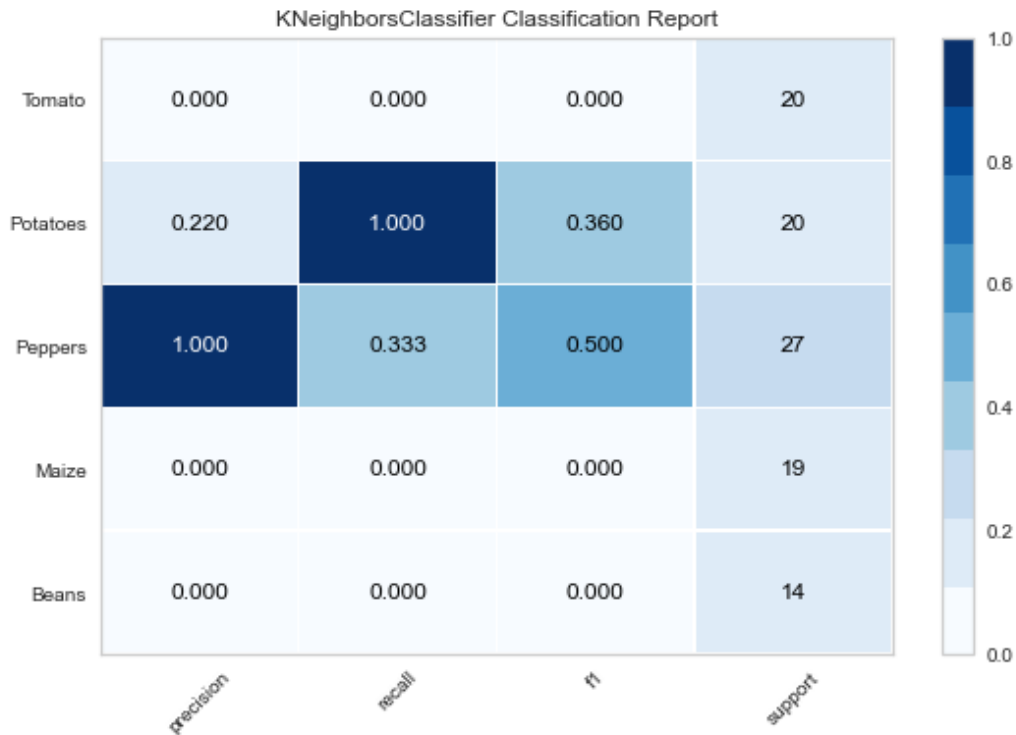


Figure 4.16 Classification report for KNN

The classification report shown in Figure 4.16 for KNN classifier shows a precision of 22% for potatoes and 33.3% recall for peppers. This means that for all positively predicted potatoes only 22% were correctly predicted. For all the positive peppers instances predicted only 33.3% were correctly classified.

The zero-figure shows that there is a problem with the classification. This classification report is done on the data that has not been standardised. Figure 4.15 shows the results for the dataset that was scaled. The report has no zeros and shows a 100% precision and recall.

iii. Classification using Random Forest Classifier

Figure 4.17 shows the accuracy score of 100% on both the train and test set for the Random Forest classifier meaning that the model can make the prediction with an accuracy of 100%.

Figure 4.18 shows the classification report for the Random Forest classifier. The report shows the precision and recall of 100% for all classes. This means that for all the positive class predictions, all were correct, meaning that 100% of all the positive class predictions all of them were correctly classified.

Both the accuracy score and the classification report are in conformity on their results. It is important to evaluate using both the classification report and the confusion matrix because classification accuracy score does not provide the level of detail required to diagnose the model's performance.

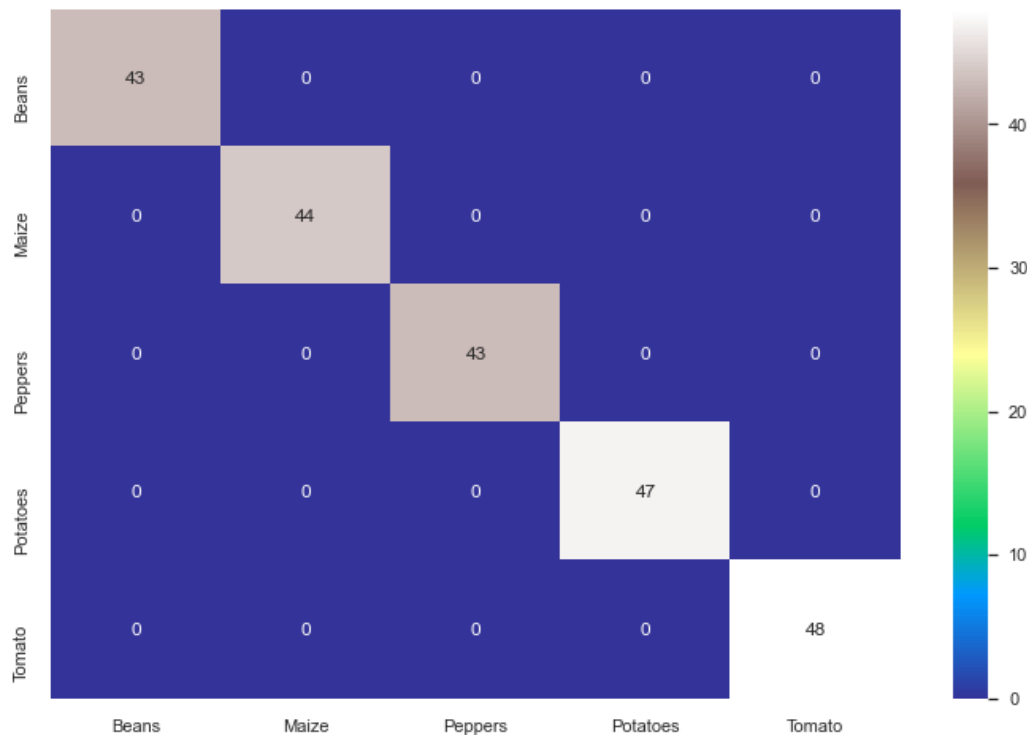


Figure 4.17 Confusion matrix for Random Forest classifier



Figure 4.18 Classification report for Random Forest classifier

iv. Classification using Gradient Boosting Classifier

Figure 4.20 depicts the prediction results for the Gradient Boosting classifier. The accuracy score according to the results is 100%. This means that the classifier is able to make predictions with 100% accuracy.

Figure 4.19 is a heat map confusion matrix for the Gradient Boosting classifier. The value in the diagonal line shows the number of times the classifier predicted the actual class correctly. The values outside of the diagonal line shows the number of times the classifier made an error. For example, the second block in the top row determines how often the model actually predicted beans as maize. This means that the model is not confusing the classes or rather it is not getting confused with the features used to classify.

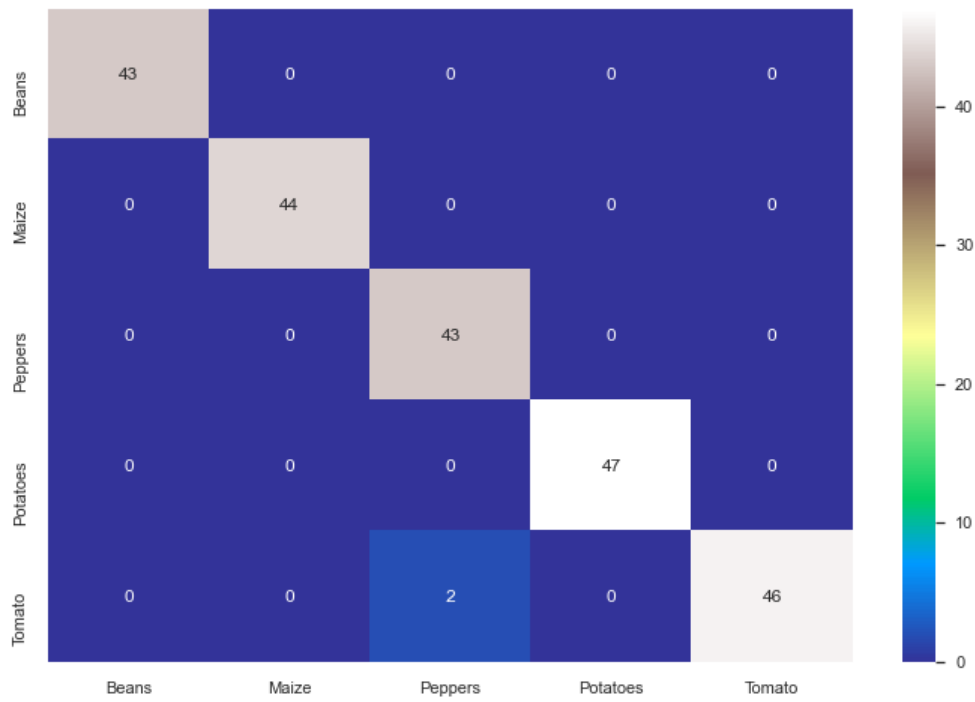


Figure 4.19 Confusion matrix for the Gradient Boosting classifier

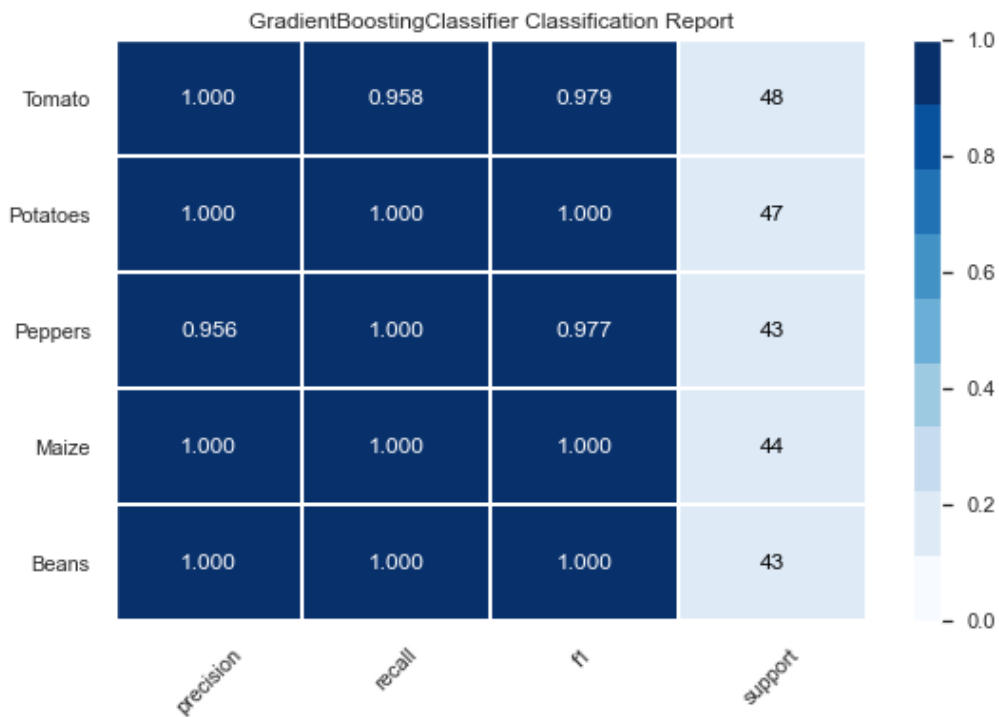


Figure 4.20 Classification report for Gradient Boosting classifier

v. Classification using Support Vector Machine Classifier (SVM)

Figures that follow below illustrate the results for Support Vector Machine classifier for linear, Radial bias function (RBF) and Polynomial kernel respectively. Kernel functions are used because intricate mathematical functions are used with SVM. A kernel function is a way of taking input data and transforming it into processing format required. Kernel functions are implemented by using “scikit-learn” library. The results indicate a 99.5% accuracy score for the linear kernel, meaning the kernel can make a prediction with 99.5% accuracy.

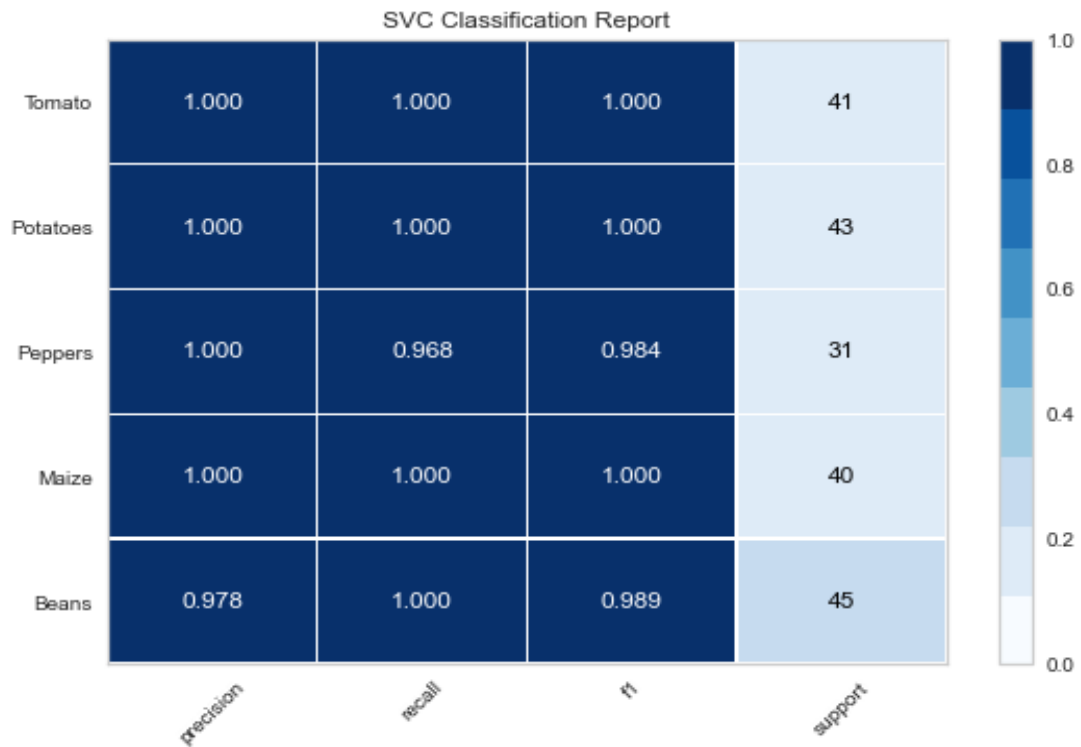


Figure 4.21 Classification Report for SVM linear kernel

Figure 4.22 displays the confusion matrix for the SVM linear kernel classifier. The classifier erroneously predicted peppers as beans one time. The classification report shows a precision of 97.8% for beans class and a recall of 96.8% for the pepper class. This means that the model experienced uncertainty between the two classes.

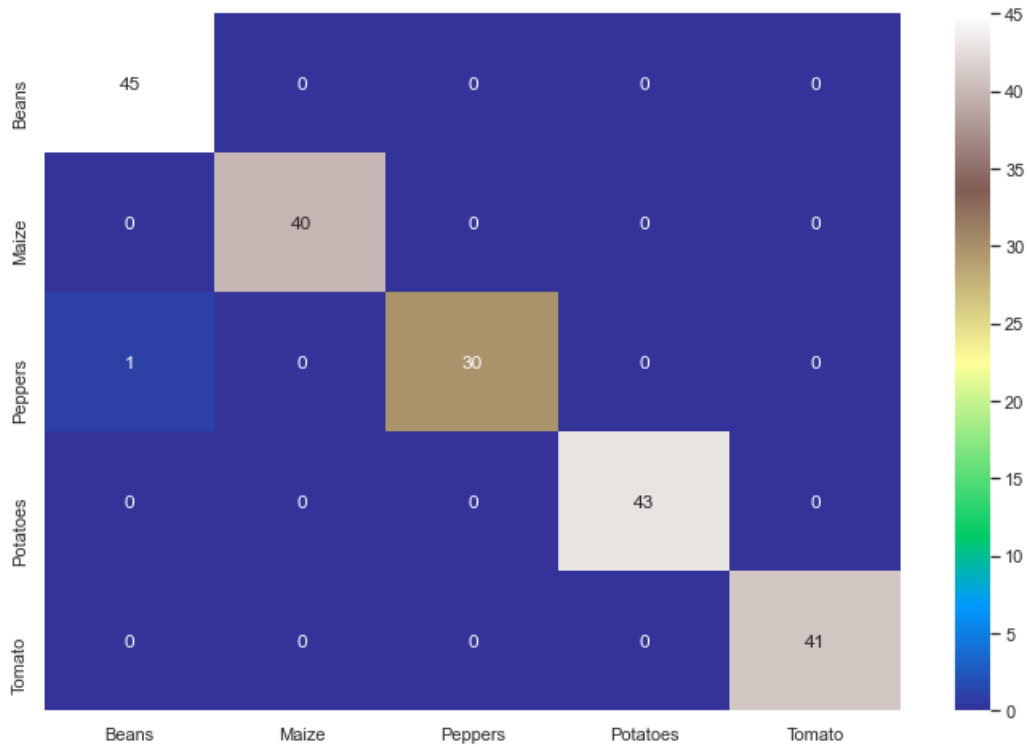


Figure 4.22 Confusion matrix for the SVM linear kernel

Figures 4.24 and 4.23 renders a representation of the Confusion matrix and Classification report for the SVM Radial bias function classifier. The model produced an accuracy score of 100%. The Precision and Recall according to Figure 4.24 is 100%. The confusion matrix shows no confusion with regards to classifying the instances.

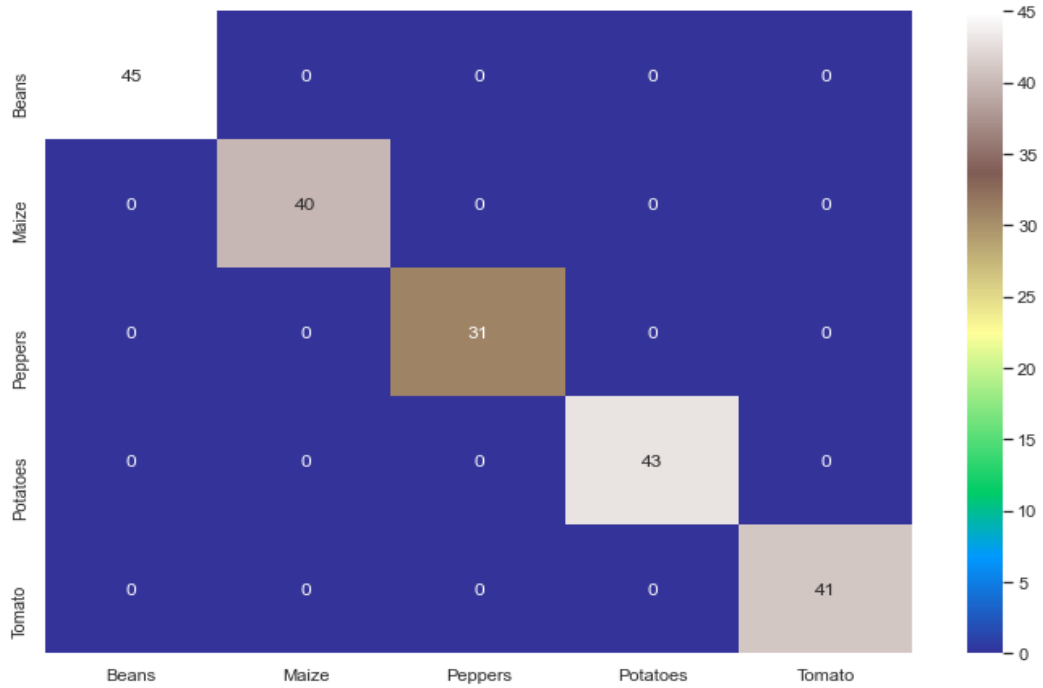


Figure 4.23 Confusion matrix for SVM Radial bias function classifier

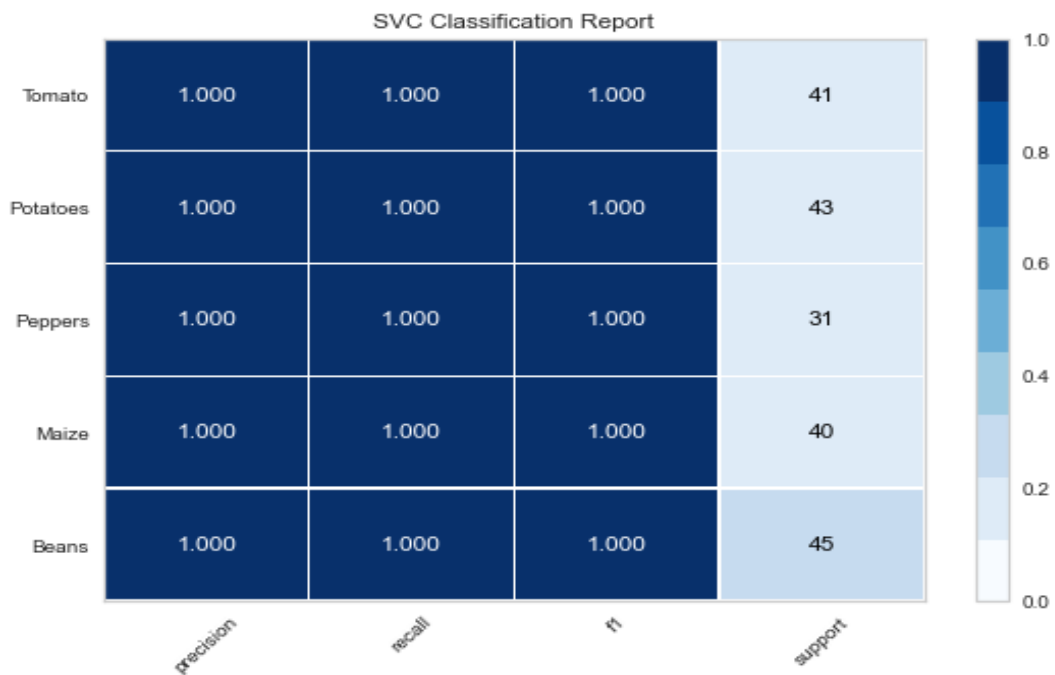


Figure 4.24 Classification report for SVM Radial bias classifier

Figures 4.25 and 4.26 portray a Confusion matrix and Classification respectively for the SVM Polynomial kernel classifier. As seen earlier, the classifier depicted an accuracy

score of 100% for this kernel. The Confusion matrix and the Classification report are concurring with those results. The confusion matrix delineates no confusion in classification and the classification report reveals a Precision and Recall of 100%.

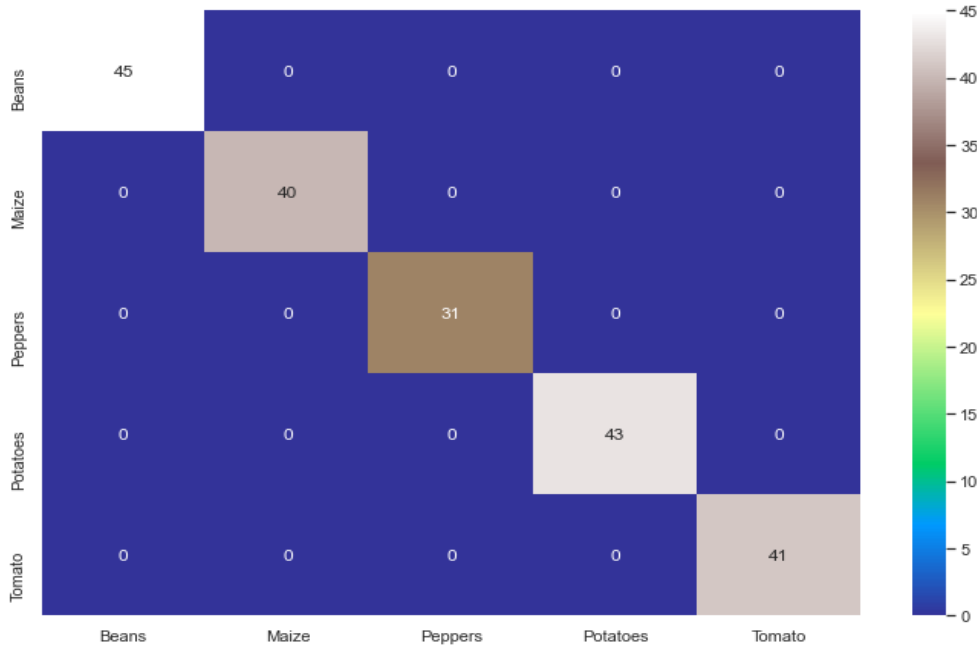


Figure 4.25 Confusion matrix for SVM Polynomial classifier

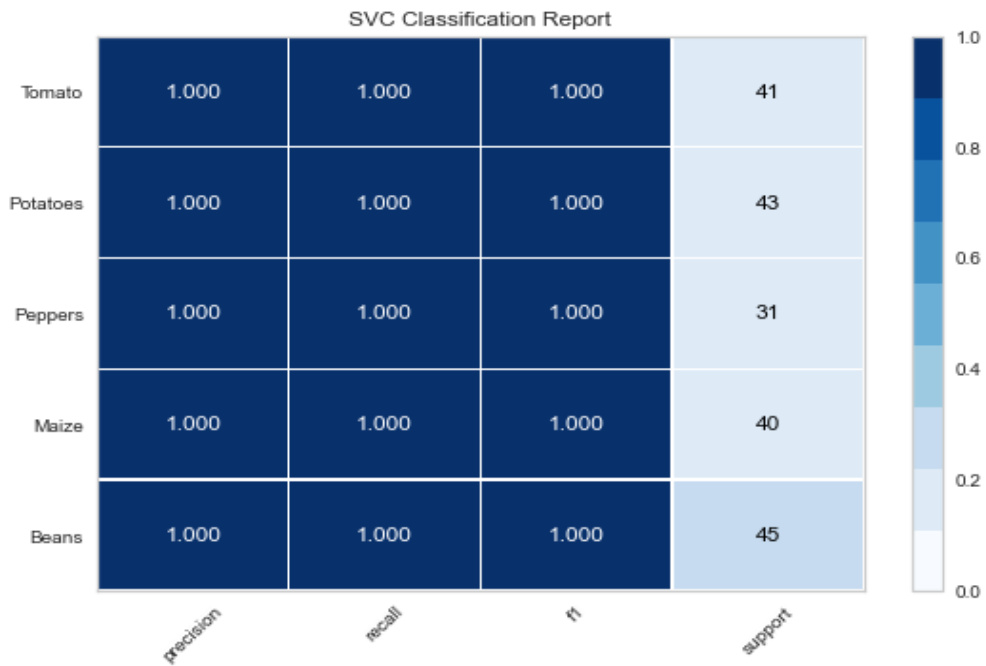


Figure 4.26 Classification report for SVM polynomial classifier

4.3 Findings

As expected, the algorithms that required data scaling produced results that were misleading and were misrepresentative of the expected results when applied to data that was not scaled. Out of all the edaphic features that were exposed to the ML algorithms, only one was taken into consideration in the decision's outcome except for Decision trees, which took both Nitrogen and Phosphorous into consideration. As a result, it is reasonable to assume that climatic factors influence crop selection decisions the most. It can also be concluded that machine learning can be employed in precision agriculture to predict the crop selection prior to planting under particular climatic and edaphic conditions.

4.4 Conclusion

In this chapter, a multi-class classification was evaluated using predictive modelling by assessing classification algorithms available for multi-class classification problems. It is crucial to understand how the model will function with data that has not been seen before. As a result, assessing the model's performance is critical.

To prevent overfitting, the evaluation may not be done on same data used for training. Evaluation of performance may be done in one of two ways namely: predictions may be done on data that you already have answers to or by using statistical techniques called resampling methods. There are four types of resampling methods used for prediction or for calculating the accuracy. Train and Test sets, k-fold Cross Validation, Leave One Out Cross Validation, Repeated Random Test-Train Splits. The results for the accuracy score as shown in the Figures were obtained by randomly splitting the data into Train and Test sets method.

The performance metrics used are Classification accuracy, Classification report and Confusion matrix. The plots for the confusion metrics and classification reports are shown in the number of figures shown in the chapter. Predictive modelling, which employs historical data to develop predictions about current data, is the most valuable aspect of machine learning. The focus on improving predicted findings is constantly present.

Data, Algorithms, Tuning, and Ensembles can all be used to improve performance. Because the data was collected using brute force in this research, the enhancement on promotion was done with the data and the Algorithms. Since the data was created, there was no data sanitisation required. Comparison of the results and possible performance enhancement will be discussed in the next chapter.

Chapter 5: Simulation results and discussion

5.1 Introduction

The objective of this chapter is to discuss the results of the Machine learning algorithms employed in Chapter 4 to develop a model for crop selection, using the multi-class classification in Python Jupyter Notebook from the Anaconda package. The purpose is to establish which is the best model for the associated problem and whether any further enhancements are possible. In the previous chapter, five different models were developed using the crop selection data that was generated using brute force. The results of the evaluation of these models will be put under microscope and some of the motives for specific actions will be examined. The comparison of the various models that were utilised will also be scrutinised. Different metrics will be exploited in the comparison of the distinct models to get to the best case.

As already mentioned, performance of a model may be refined by deploying different techniques. Some of these techniques will be revisited in this chapter and discussed further. The multi-class classification models reviewed in the previous chapter include the three Ensemble classifiers, which is Decision tree classifier, Random Forest classifier, and Gradient boosting classifier. Furthermore, the K-Nearest Neighbours and the Support Vector Machines were also considered. The results of the evaluations follow next.

5.2 Simulations results

5.2.1 Accuracy score

Classification Accuracy score is the measure of the number of correct predictions in the given data sample and it is measured by dividing the number of correct predictions with the total number of instances in the sample. Accuracy score works very well when there is an equal number of instances of the different classes in a sample. Table 5.1 reflects the accuracy scores for the different models as encountered in the previous chapter.

The goal of the Ensemble classifier is to improve robustness over a single estimator, aggregate the predictions of numerous base estimators created with a particular learning technique. This is achieved by averaging or boosting methods. In averaging methods, the principle is to construct numerous estimators individually before averaging their forecasts. In boosting, base estimators are built in a sequential order, with the aim of reducing the aggregate estimator's bias. The Decision Tree and the Random Forest (RF) use averaging method, and the Gradient boosting uses the boosting method. From Table 5.1 it is apparent that the Random Forest classifier presents a better accuracy score than that of the Decision Tree classifier. The accuracy score for the Random Forest classifier is 100%, and that of the Decision Tree classifier is 99.1%. The Gradient Boosting (GB) classifier produced the same accuracy score as the Decision tree classifier (99.1%). This is not surprising as the base for the GB is decision trees which run sequentially with corrective measures from one tree to the next. GB is susceptible to overfitting if the data has a lot of noise. GB could only achieve an accuracy score of 100% if unlimited depth of trees are used in which case the algorithm will memorise rather than train the data. On the contrary, the RF is less likely to overfit but much easier to tune. RF uses multiple decision trees and therefore can achieve the 100% accuracy for the reasons mentioned earlier when discussing the GB.

The Support Vector Machine classifier uses statistical learning frameworks to make prediction. In the previous chapter, three functions of the SVM classifier were interrogated, namely linear function, Radial Bias Function, and the Polynomial function. As the rule of thumb, the Linear Function is considered first. The linear function obtained an accuracy score of 99.5% as indicated in Table 5.1.

The K-Nearest Neighbour selects the value of K and searches for one nearest observation. KNN classifier therefore memorises the dataset. This means that the accuracy score will always be 100% because the testing is happening on the same data. It will always make correct prediction. This proved true with the results as seen in Table 5.1, where the accuracy score for this classifier is 100%.

Table 5.1 Summary of Accuracy scores

Model name	Accuracy score
Decision Tree	0.991
Random Forest	1.00
Gradient Boosting	0.991
K-Nearest Neighbour	1.00
Support Vector Machine:	
Linear function	0.995
Radial Bias Function	1.00
Polynomial Function	1.00

5.2.2 Confusion Matrix

Because accuracy score works so well in cases where the number of instances per category is the same, it cannot be relied on to be the best measure of the performance of the model. In a paper by [57] the author alludes to the fact that accuracy alone is not to evaluate the performance of the model. For this reason, the Confusion Matrix is also used to determine how well the model can make predictions.

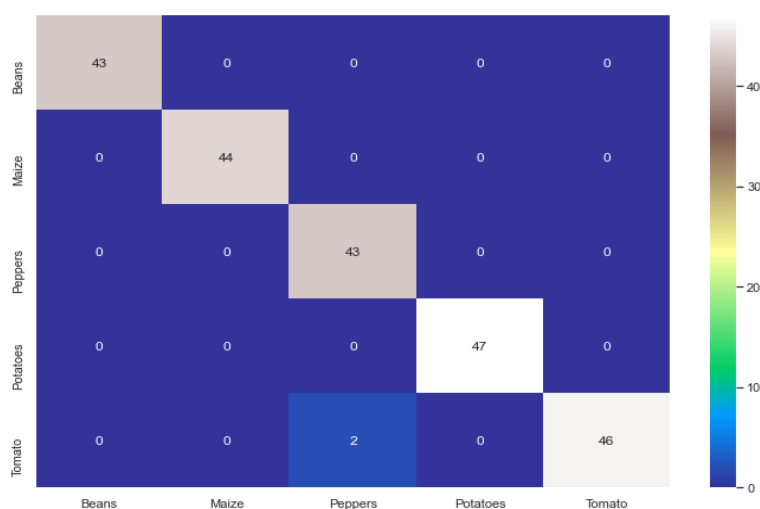


Figure 5.1 Confusion Matrix for Decision tree classifier

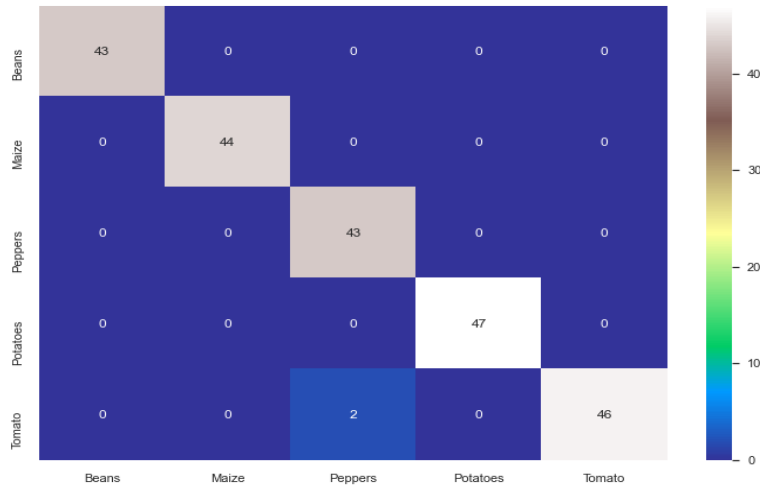


Figure 5.2 Confusion Matrix for Gradient Boosting classifier

Figures 5.1 and 5.2 exhibits the Confusion Matrix for Decision Tree classifier and the Gradient Boosting classifier respectively, with the accuracy score of 99.1% as tabled above. The Both Decision Tree classifier and the Gradient Boosting classifier predicted the actual Tomato class to be Peppers twice. Both classifiers correctly predicted the Tomato class 46 times out of 48 available instances and confused the Tomato class for Peppers twice. Both classifiers exhibited the same results in their confusion matrix.

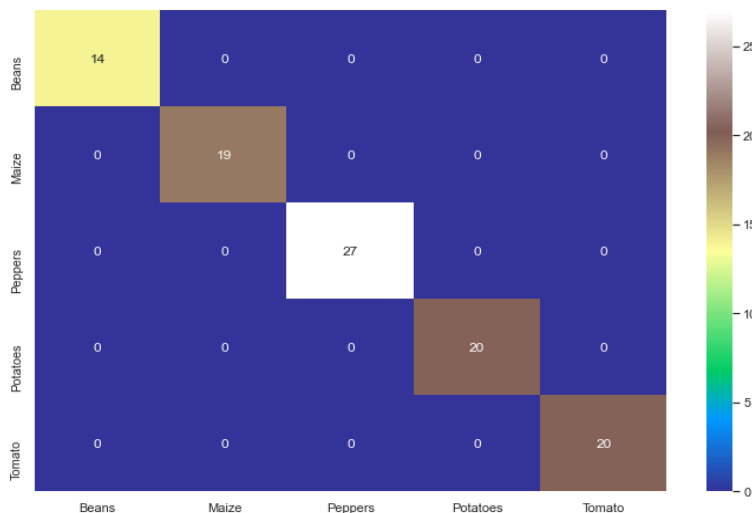


Figure 5.3 Confusion matrix for KNN classifier

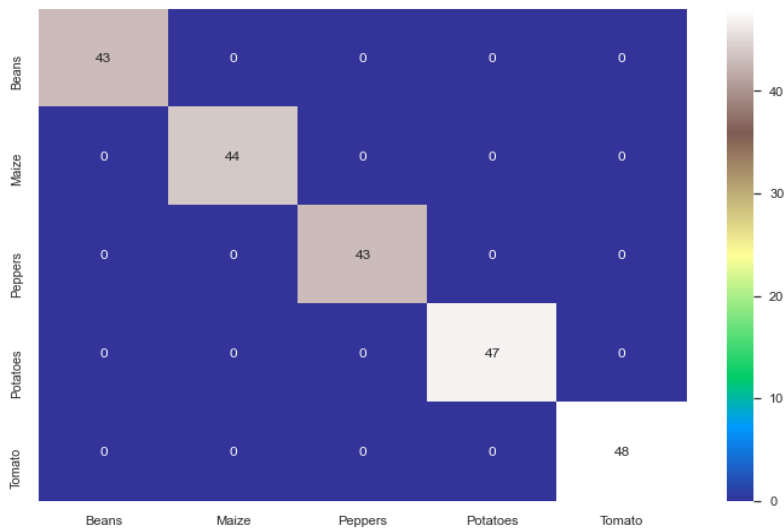


Figure 5.4 Confusion Matrix for Random Forest

Figures 5.3 and 5.4 depict Confusion matrix for the K-Nearest Neighbour and the Random Forest respectively. Both classifiers dispose no confusion in terms of classification. The classification in the said classifiers presents a diagonal line with the total number of instances classified. The classifiers classified all instances in the given sample correctly.

Figure 5.5 displays the Confusion matrix for the Support Vector Machine linear classifier. The classifier confused the actual Pepper class to be Beans one time. The classifier classified 45 of the 46 instances in the sample correctly.

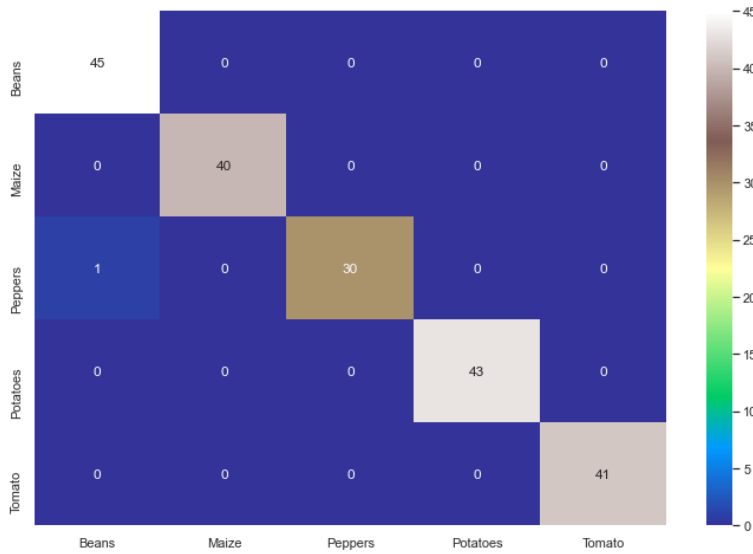


Figure 5.5 Confusion Matrix for SVM linear classifier

5.2.3 Classification report

Table 5.2 disposes a summary of the Classification reports for the different models as simulated in the previous chapter. The four metrics of evaluation for model performance review are studied closely and compare in this chapter. The metrics are Precision, Recall, f1-score and Support.

Precision, also called the Positive predictive value, is the rate of true positives over all the positive predictions in a given sample. In other words, when the model predicts positive, how often is it correct. The Decision Tree and the Gradient Boosting produced a Precision of 0.956 for the Peppers classification, meaning that when they classified Peppers, they were correct only 95.6% of the time. The support Vector Machine linear classifier obtained a Precision of 0.987 for the Beans class. This means that when the classifier made a Beans classification it was correct only 98.7% of the time.

Recall, also known as the sensitivity, is the measure of how accurately the model is able to identify relevant data. This means for all actual class predictions how many were correctly classified. For this metric there were three models who had a less than 100% value. Decision tree, Gradient Boosting and Support Vector Machine classifiers obtained 95.8%, 95.8% and 96.8% respectively.

F1-Score is the harmonic mean between the Precision and the Recall. It provides a trade-off between the two metrics. Once again, the three models with less than 100% F1-score were Decision trees, Gradient Boosting and the Support Vector Machine.

Table 5.2 Summary of classification reports

Model	Category	Precision	Recall	F1-Score	Support
Decision tree	Tomatoes	1.0	0.958	0.979	48
	Potatoes	1.0	1.0	1.0	47
	Peppers	0.956	1.0	0.977	43
	Maize	1.0	1.0	1.0	44
	Beans	1.0	1.0	1.0	43
Random Forest	Tomatoes	1.0	1.0	1.0	48
	Potatoes	1.0	1.0	1.0	47
	Peppers	1.0	1.0	1.0	43
	Maize	1.0	1.0	1.0	44
	Beans	1.0	1.0	1.0	43
Gradient Boosting	Tomatoes	1.0	0.958	0.979	48
	Potatoes	1.0	1.0	1.0	47
	Peppers	0.956	1.0	0.977	43
	Maize	1.0	1.0	1.0	44
	Beans	1.0	1.0	1.0	43
Support Vector Machine	Tomatoes	1.0	1.0	1.0	41
	Potatoes	1.0	1.0	1.0	43
	Peppers	1.0	0.968	0.984	31
	Maize	1.0	1.0	1.0	40
	Beans	0.978	1.0	0.989	45
Linear	Tomatoes	1.0	1.0	1.0	21
	Potatoes	1.0	1.0	1.0	30
	Peppers	1.0	1.0	1.0	30
	Maize	1.0	1.0	1.0	21
	Beans	1.0	1.0	1.0	23
K-Nearest Neighbour	Tomatoes	1.0	1.0	1.0	21
	Potatoes	1.0	1.0	1.0	30
	Peppers	1.0	1.0	1.0	30
	Maize	1.0	1.0	1.0	21
	Beans	1.0	1.0	1.0	23

5.3 Process of elimination

Five models have been evaluated and only one can be selected to be employed for the task. Though the K-Nearest Neighbour exhibits what seem to be the best results of all the other model that is 100% accuracy score, 100% Precision, 100% Recall and 100% F1-score, the method that it uses in making predictions is undesirable for the task. It does not try to build a generic internal model; instead, it just saves instances of the training data. Also, according to [58], 100% classification accuracy could be harmful. Therefore, this model is eliminated as the worst of all the models.

When looking at the three Ensemble classifiers, they can be compared with one another because they all use the same process for classification. They all use an aggregate of multiple classifiers to get the best result. The Decision tree and the Gradient Boosting obtained the same results in the Accuracy score and their confusion was in the same categories, though the one works with averaging and the other with boosting. Therefore, the Random Forest selects the best of the three because it performed at 100% Accuracy, Precision, Recall and F1-score. The Random Forest suffered no confusion in classifying. The 100% classification is acceptable in this case because it performed better against classifiers using the same method, and same data.

The two classifiers that may be compared for consideration are therefore the Random Forest and the Support Vector Machine classifiers. The Random Forest model exhibits the best Accuracy score, Precision, Recall and F1-score. However, according to the machine learning model selection process flow, the Ensemble classifiers should be considered as the last option. The Support Vector Machine Linear classifier should be considered first. According to the task at hand, it is important to predict the crop to be planted with a higher Recall than Precision, and therefore the Support Vector Machine showed the highest Recall when compared to the other classifiers without 100% Recall. The Support Vector Machine Linear model is therefore the best model to use for the crop selection DSS.

5.4 Conclusion

In this chapter the performances of the models that were developed in the previous chapter according to the different metrics were evaluated. The Decision tree, Random Forest, Gradient Boosting, K-Nearest neighbour, and the Support Vector Machine classifiers were evaluated and compared to determine the best model for crop selection prior to planting. The metrics used were Accuracy score, Precision, Recall, F1-score and support. After all considerations, it has been concluded that the Support Vector Machine Linear classifier was found to be the best classifier for the task, according to the reasons already discussed in this chapter. This concludes the discussion about the classification model employed.

Chapter 6: Conclusion and future work

6.1 Conclusion

This chapter serves to provide a conclusion to the discussion on the solution provided by the Intelligent Agro-Climate Decision Support System (IACDSS) for small-scale farmers, and to present the model of the system. This is a summary of what the study has achieved.

6.1.1 Summary of the study

In Chapter 1, a background of the study, the problem statement, and the objectives of the study are outlined. The objectives which included data collection, analysis and the development of an intelligent agro-climate decision support system were listed.

In Chapter 2, a literature review of the technologies researched and those that are available were discussed and comparison of the solutions was tabled. Chapter 2 concluded that the available solutions were not user friendly for the small-scale farmer, and that there was a need for evaluating an amalgamation of different technologies to provide a solution for small-scale farmers.

In Chapter 3, the methodology followed to develop the said decision support system was discussed, and all data collection methods used were tabled. All data that was collected, tabled and discussed. How the different elements of the system were put together, which is, the Indigenous knowledge, mobile phone, and machine learning was explained. The framework of the study was explained in this chapter. The first objective of this study was introduced in this chapter. The purpose was to collect Indigenous knowledge, edaphic data, climatic data and demographic data of the selected area in KwaZulu-Natal. The data was collected and analysed, using the different methods that have already been mentioned in this chapter in order to draw particular conclusions that would form the basis of the successor chapter.

In Chapter 4, the second objective of this study was introduced. In order to determine whether the subdivision of the AI that is, ML could be explored to provide the solution to crop selection decision prior to planting, ML were interrogated. Classification using the different algorithms was utilised and evaluated using different metrics to provide the best possible model for the solution. It was determined that classification using multiple classes of crops and climatic and edaphic attributes was possible, and that all possible algorithms to use performed satisfactorily. The next chapter introduces the process elimination to select the best algorithm.

The performance of the algorithms was discussed in Chapter 5 and the process of elimination was undertaken to select the best algorithm from support vector machine linear algorithm, gradient boosting, random forest and the decision tree algorithms. The model performance was evaluated using the different metrics to determine supreme performer. The metrics used were Accuracy score, Precision, Recall, and the F1-score. The procedure revealed that the Support Vector Machine Linear classifier was the top performing model to use for classification of crop selection at planting stage, with the accuracy score of 99.5%. This draws to a close another objective of this study which was to ascertain by how much the ML would improve the level of accuracy of the small-scale farmer's decision at planting stage. Crop selection at planting stage forms part of the tactical decisions that small-scale farmers need to take as part of agricultural decision management. The crop selection can be predicted with an accuracy of 99.5%.

Finally, this chapter concludes the objectives of this study, which is to develop an integrated decision support system for small-scale farmers. In Chapter 2 and 3, Indigenous knowledge was discussed as part of the solution, and how it was implemented in this particular case. As discussed already, the IACDSS uses the framework or follows the blueprint of the ITIKI model. Chapter 3 explains the methodology used to develop the IACDSS model. The ITIKI solution combines climatic conditions with the Indigenous knowledge and uses fuzzy logic to make drought prediction. In the IACDSS, the edaphic conditions that were utilised for the simulation done in Chapter 4 represents the agricultural Smart Sensors information.

In Chapter 4, the Machine learning model was developed and trained using different algorithms and comparing them to find the best suitable one for selection of crops prior to planting. Section 6.1.2 provides a summary of the IACDSS's functionality.

6.1.2 Summary of the functionality of the IACDSS

The IACDSS model as show in Figure 6.1, takes edaphic data (b) and climatic (a) data and runs it through ML system to predict the crop to be selected (x) for planting. The IK as observed will predict the crop to be planted according to the knowledge base and give the prediction (y). The system compares the IK decision (y) prediction and the ML prediction (x) to provide the integrated solution and disseminate the information to the farmer if the IK prediction is the same as the ML prediction ($x=y$). In the case where the IK decision prediction and the ML prediction is not the same ($x\neq y$), then the two go to reconciliation where a decision would either be intercropping or a decision by an expert crop scientist. When a decision has been found, the information will be disseminated using mobile phones, by means of short messaging system (SMS) or push notification in the application itself.

It is therefore evident that machine learning can be applied to agro-climate conditions in agriculture to assist in decision support. This study has shown evidence of mobile phone technology being exploited as a collaborative tool for solution delivery, case in point the ITIKI framework. The said framework has already provided evidence that Indigenous knowledge can be used in combination with scientific models to make predictions for rainfall with success. This forms the basis of the Intelligent Agro-climate Decision support tool. Rainfall predictions precedes the decision of the type of crop to be planted. Therefore, since an Intelligent Agro-climate Decision Support tool for small-scale farmers proceeds a successfully implemented ITIKI system, it can be expected that the achieved solution will have the same success.

In conclusion, the objectives of this study as already mentioned in Chapter 1 were data collection and analysis, developing a framework for integration of IK and ML for cropping decisions, and determining how much the level of accuracy of decisions would be impacted by the given solution. It can therefore be concluded that the first two

objectives were addressed in this study, as discussed in Chapter 3. It further determined that ML model can predict the selection of the crop to be planted with an accuracy of 99.5% according to the given condition as discussed in Chapter 5. It can therefore be concluded that the three research aims in this study have been addressed.

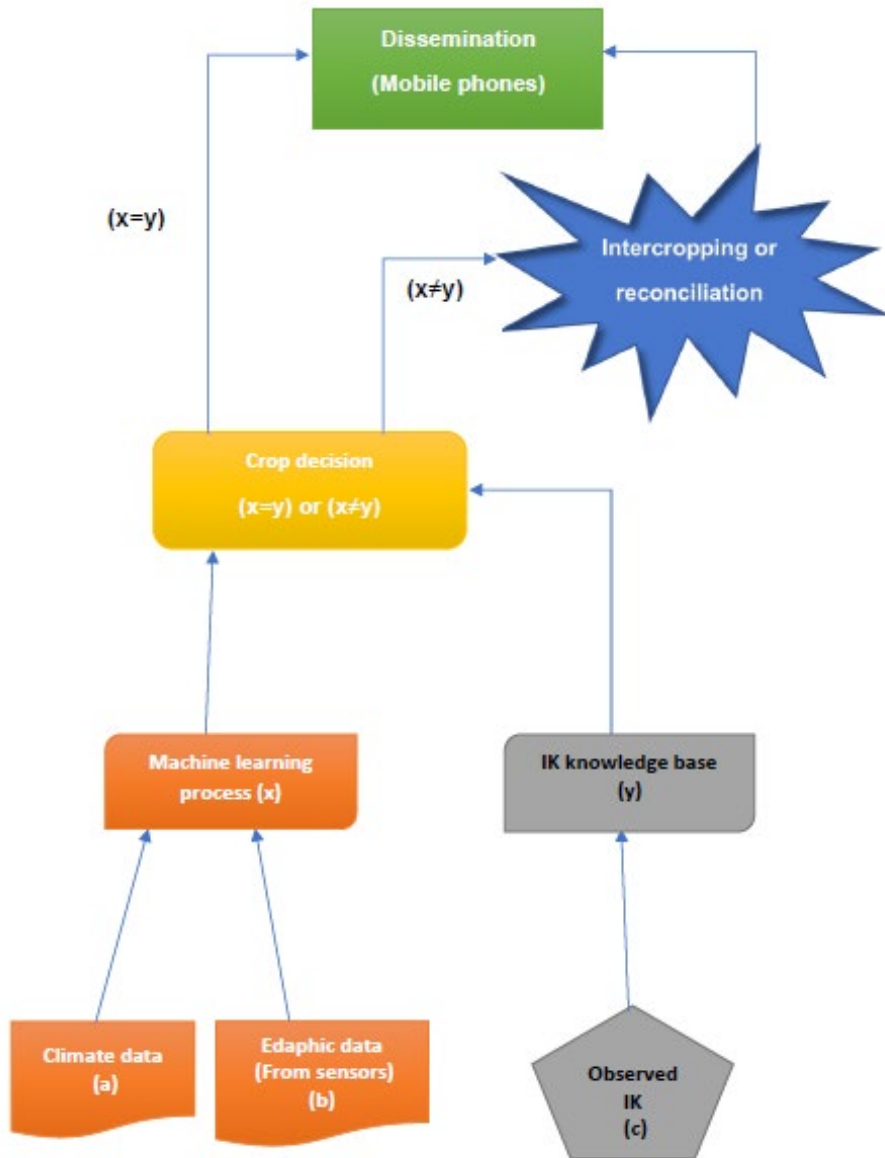


Figure 6.1 The Intelligent Agro-Climate Decision Support System representation

6.2 Future work

Given the restriction imposed by the Covid-19 Pandemic, the Agro-Climate Decision Support System was not tested in the field under real-life situations. For further work therefore, it is proposed that the system be deployed in the field study (in Swayimane) and real-life data on its performance collected for at least 2 cropping seasons (equivalent to one calendar year). Both qualitative and quantitative data analysis should then be carried out to determine the extent to which it improves the livelihood of the small-scale farmers. Parameters such as yield increase and profitability could be adopted [58].

A further recommendation is to explore multi-label classification to determine intercropping options for small-scale farmers. This is for two reasons, one, to further build the framework presented in Figure 6.1 and secondly, to cater for the fact that small-scale farmers do not practice niche agriculture and have the need to plant more than one crop most times.

References

- [1] O. Westermann, W. Förch, P. Thornton, J. Körner, L. Cramer, and B. Campbell, “Scaling up agricultural interventions: Case studies of climate-smart agriculture,” *Agric. Syst.*, vol. 165, no. June, pp. 283–293, 2018, doi: 10.1016/j.agsy.2018.07.007.
- [2] IFRC, *World Disasters Report 2020: Come Heat or High Water*. 2020. [Online]. Available: https://www.ifrc.org/sites/default/files/2021-05/20201116_WorldDisasters_Full.pdf
- [3] GDO analytical Report, “GDO Analytical Report GDO Analytical Report Drought in Sub-Saharan and Southern Africa – August,” no. August, p. 2, 2021.
- [4] M. M. Akanbi A.K, “Towards the Development of a Rule-Based Drought Early Warning Expert Systems Using Indigenous Knowledge,” *Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD), Durban, 2018*, pp. 1–8, 2018, doi: 10.1109/ICABCD.2018.8465465.
- [5] G. Amarnath *et al.*, “Using smart ICT to provide weather and water information to smallholders in Africa: The case of the Gash River Basin, Sudan,” *Clim. Risk Manag.*, vol. 22, no. September 2016, pp. 52–66, 2018, doi: 10.1016/j.crm.2018.10.001.
- [6] J. F. Kirsten and J. Van Zyl, “Defining small-scale farmers in the south african context,” *Agrekon*, vol. 37, no. 4, pp. 551–562, 1998, doi: 10.1080/03031853.1998.9523530.
- [7] M. Masinde, “MAS-DEWS: A multi-agent system for predicting Africa’s drought,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2015-Septe, 2015, doi: 10.1109/IJCNN.2015.7280837.

- [8] M. Masinde, “IoT applications that work for the African continent: Innovation or adoption?,” *Proc. - 2014 12th IEEE Int. Conf. Ind. Informatics, INDIN 2014*, pp. 633–638, 2014, doi: 10.1109/INDIN.2014.6945587.
- [9] M. Masinde, “An innovative drought early warning system for sub-saharan Africa: Integrating modern and indigenous approaches,” *African J. Sci. Technol. Innov. Dev.*, vol. 7, no. 1, pp. 8–25, 2015, doi: 10.1080/20421338.2014.971558.
- [10] M. Masinde and P. N. Thothela, “ITIKI Plus: A Mobile Based Application for Integrating Indigenous Knowledge and Scientific Agro-Climate Decision Support for Africa’s Small-Scale Farmers,” *2019 IEEE 2nd Int. Conf. Inf. Comput. Technol. ICICT 2019*, pp. 303–309, May 2019, doi: 10.1109/INFOCT.2019.8711059.
- [11] M. Masinde, A. Bagula, and N. Muthama, “Implementation roadmap for downscaling drought forecasts in Mbeere using ITIKI,” *Int. Telecommun. Union - Proc. 2013 ITU Kaleidosc. Acad. Conf. Build. Sustain. Communities, K 2013*, pp. 63–70, 2013, [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84881272297&partnerID=tZOtx3y1>
- [12] R. Kabbiri, M. Dora, V. Kumar, G. Elepu, and X. Gellynck, “Mobile phone adoption in agri-food sector: Are farmers in Sub-Saharan Africa connected?,” *Technol. Forecast. Soc. Change*, vol. 131, no. December 2016, pp. 253–261, 2018, doi: 10.1016/j.techfore.2017.12.010.
- [13] S. Adepun and R. F. Adler, “A comparison of performance and preference on mobile devices vs. desktop computers,” *2016 IEEE 7th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2016*, 2016, doi: 10.1109/UEMCON.2016.7777808.

- [14] M. Toseef and M. J. Khan, “An intelligent mobile application for diagnosis of crop diseases in Pakistan using fuzzy inference system,” *Comput. Electron. Agric.*, vol. 153, no. August, pp. 1–11, 2018, doi: 10.1016/j.compag.2018.07.034.
- [15] Food and Agriculture Organization of the United Nations, *Food Security and Nutrition in the World Security, Improved Nutrition and Affordable Healthy Diets for All*. 2021.
- [16] N. Jain, A. Kumar, S. Garud, V. Pradhan, and P. Kulkarni, “Crop Selection Method Based on Various Environmental Factors Using Machine Learning,” *Int. Res. J. Eng. Technol.*, vol. 4, no. 2, pp. 1530–1533, 2017, [Online]. Available: <https://irjet.net/archives/V4/i2/IRJET-V4I2299.pdf>
- [17] E. Han *et al.*, “SIMAGRI: An agro-climate decision support tool,” *Comput. Electron. Agric.*, no. April, pp. 0–1, 2018, doi: 10.1016/j.compag.2018.06.034.
- [18] P. Sondhi, “Feature construction methods: a survey,” *Sifaka. Cs. Uiuc. Edu*, vol. 69, pp. 70–71, 2010.
- [19] Y. Vunnam, P. T, and K. R. Vunnam, “Evolution of Mobile Wireless Communication,” *Int. J. Recent Trends Eng. Res.*, vol. 05, no. 11, pp. 1–9, 2019, doi: 10.23883/ijrter.2019.5089.udx4d.
- [20] V. F. Jeffers, F. Humber, T. Nohasiarivelo, R. Botosoamananto, and L. G. Anderson, “Trialling the use of smartphones as a tool to address gaps in small-scale fisheries catch data in southwest Madagascar,” *Mar. Policy*, vol. 99, no. May 2018, pp. 267–274, 2019, doi: 10.1016/j.marpol.2018.10.040.
- [21] C. Z. Qiang, S. C. Kuek, A. Dymond, and S. Esselaar, “Mobile Applications for Agriculture and Rural Development,” no. December, 2011.
- [22] E. Beza, P. Reidsma, P. M. Poortvliet, M. M. Belay, B. S. Bijen, and L. Kooistra, “Exploring farmers’ intentions to adopt mobile Short Message Service (SMS) for

- citizen science in agriculture,” *Comput. Electron. Agric.*, vol. 151, no. February 2017, pp. 295–310, 2018, doi: 10.1016/j.compag.2018.06.015.
- [23] S. S. Meijer, D. Catacutan, O. C. Ajayi, G. W. Sileshi, and M. Nieuwenhuis, “The role of knowledge, attitudes and perceptions in the uptake of agricultural and agroforestry innovations among smallholder farmers in sub-Saharan Africa,” *Int. J. Agric. Sustain.*, vol. 13, no. 1, pp. 40–54, 2015, doi: 10.1080/14735903.2014.912493.
- [24] D. C. Phelan *et al.*, “Advancing a farmer decision support tool for agronomic decisions on rainfed and irrigated wheat cropping in Tasmania,” *Agric. Syst.*, vol. 167, no. September, pp. 113–124, 2018, doi: 10.1016/j.agry.2018.09.003.
- [25] “DrumNet - Kenya | The Communication Initiative Network.” <https://www.comminit.com/content/drumnet-kenya> (accessed Sep. 25, 2021).
- [26] A. Manjula and G. Narsimha, “XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction,” *Proc. 2015 IEEE 9th Int. Conf. Intell. Syst. Control. ISCO 2015*, 2015, doi: 10.1109/ISCO.2015.7282311.
- [27] T. S. Jayne *et al.*, “Africa’s changing farm size distribution patterns: the rise of medium-scale farms,” *Agric. Econ. (United Kingdom)*, vol. 47, pp. 197–214, 2016, doi: 10.1111/agec.12308.
- [28] U. Gneiting and J. Sonenshine, “A LIVING INCOME FOR SMALL-SCALE FARMERS Tackling unequal risks and market power,” *Oxfam Discuss. Pap.*, no. November, 2018, [Online]. Available: www.oxfam.org
- [29] M. Meroni, F. Waldner, L. Seguini, H. Kerdiles, and F. Rembold, “Yield forecasting with machine learning and small data: What gains for grains?,” *Agric. For. Meteorol.*, vol. 308–309, no. April, 2021, doi: 10.1016/j.agrformet.2021.108555.

- [30] W. Fan, C. Chong, G. Xiaoling, Y. Hua, and W. Juyun, "Prediction of Crop Yield Using Big Data," *2015 8th Int. Symp. Comput. Intell. Des.*, pp. 255–260, 2015, doi: 10.1109/ISCID.2015.191.
- [31] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, pp. 57–65, 2016, doi: 10.1016/j.compag.2015.11.018.
- [32] N. Gandhi, O. Petkar, and L. J. Armstrong, "Rice crop yield prediction using Artificial Neural Networks," *IEEE Int. Conf. Technol. Innov. ICT Agric. Rural Dev.*, no. Tiar, pp. 105–110, 2016.
- [33] T. Nyabako, B. M. Mvumi, T. Stathers, S. Mlambo, and M. Mubayiwa, "Predicting *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: A machine learning approach," *J. Stored Prod. Res.*, vol. 87, p. 101592, 2020, doi: 10.1016/j.jspr.2020.101592.
- [34] L. Li, S. Zhang, and B. Wang, "Plant Disease Detection and Classification by Deep Learning - A Review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.
- [35] M. Á. Rodríguez-García, F. García-Sánchez, and R. Valencia-García, "Knowledge-based system for crop pests and diseases recognition," *Electronics (Switzerland)*, vol. 10, no. 8, 2021. doi: 10.3390/electronics10080905.
- [36] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors (Switzerland)*, vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [37] S. Šūmane *et al.*, "Local and farmers' knowledge matters! How integrating

- informal and formal knowledge enhances sustainable and resilient agriculture,” *J. Rural Stud.*, vol. 59, pp. 232–241, 2018, doi: 10.1016/j.jrurstud.2017.01.020.
- [38] A. T. M. S. Ahamed *et al.*, “Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh,” *2015 IEEE/ACIS 16th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2015 - Proc.*, 2015, doi: 10.1109/SNPD.2015.7176185.
- [39] R. D. Baruah, S. Roy, R. M. Bhagat, and L. N. Sethi, “Use of data mining technique for prediction of tea yield in the face of climate change of Assam, India,” *Proc. - 2016 15th Int. Conf. Inf. Technol. ICIT 2016*, pp. 0–4, 2017, doi: 10.1109/ICIT.2016.44.
- [40] T. P. Chibarabada, A. T. Modi, and T. Mabhaudhi, “Nutrient content and nutritional water productivity of selected grain legumes in response to production environment,” *Int. J. Environ. Res. Public Health*, vol. 14, no. 11, 2017, doi: 10.3390/ijerph14111300.
- [41] V. G. P. Chimonyo, A. T. Modi, and T. Mabhaudhi, “Assessment of sorghum–cowpea intercrop system under water-limited conditions using a decision support tool,” *Water SA*, vol. 42, no. 2, pp. 316–327, 2016, doi: 10.4314/wsa.v42i2.15.
- [42] V. G. P. Chimonyo, A. T. Modi, and T. Mabhaudhi, “Water use and productivity of a sorghum-cowpea-bottle gourd intercrop system,” *Agric. Water Manag.*, vol. 165, pp. 82–96, 2016, doi: 10.1016/j.agwat.2015.11.014.
- [43] C. Jahel, X. Augusseau, and D. Lo Seen, “Modelling cropping plan strategies: What decision margin for farmers in Burkina Faso?,” *Agric. Syst.*, vol. 167, no. August, pp. 17–33, 2018, doi: 10.1016/j.agsy.2018.08.004.
- [44] APRO Software, “Why Machine Learning Is Agriculture’s New Best Friend -

- APRO Software.” <https://apro-software.com/machine-learning-agriculture/> (accessed Sep. 25, 2021).
- [45] A. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” *Comput. Electron. Agric.*, vol. 151, no. June, pp. 61–69, 2018, doi: 10.1016/j.compag.2018.05.012.
- [46] R. S. Sicut, E. J. M. Carranza, and U. B. Nidumolu, “Fuzzy modeling of farmers’ knowledge for land suitability classification,” *Agric. Syst.*, vol. 83, no. 1, pp. 49–75, 2005, doi: 10.1016/j.agsy.2004.03.002.
- [47] Masinde Euphraith Muthoni, “ITIKI: Bridge between African Indigenous Knowledge and Modern Science on Drought Prediction,” UNIVERSITY OF CAPE TOWN, 2012. doi: 10.1080/19474199.2012.683444.
- [48] P. S. Vijayabaskar, “Crop Prediction Using Predictive Analytics,” pp. 370–373, 2017.
- [49] A. Bryman, E. Bell, P. Hirschsohn, A. dos Santos, J. du Toit, and A. Masenge, “Research Methodology: Business and Management Contexts,” *Oxford University Press*, 2014.
- [50] R. J. Sternberg, “Intelligence,” *Dialogues Clin. Neurosci.*, vol. 14, no. 1, pp. 19–27, 2012, doi: 10.7551/mitpress/9780262062749.003.0018.
- [51] S. for Africa, “Kwazulu Natal Vegetable Planting Chart – Seeds for Africa.” <https://www.seedsforafrica.co.za/pages/kwazulu-natal-vegetable-planting-chart> (accessed Aug. 17, 2021).
- [52] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 13001–13008, 2020, doi: 10.1609/aaai.v34i07.7000.

- [53] DALRRD, “Department of Agriculture, Land Reform and Rural Development > Branches > Agricultural Production, Health & Food Safety > Plant Production > Production Guidelines > Pguidelinesarchive.”
<https://www.dalrrd.gov.za/Branches/Agricultural-Production-Health-Food-Safety/Plant-Production/Production-Guidelines/Pguidelinesarchive> (accessed Aug. 17, 2021).
- [54] A. Pozdnoukhov and M. Kanevski, “Machine learning algorithms for analysis and modeling of geospatial data,” *Geomathematics GIS Anal. Resour. Environ. Hazards - Annu. Conf. Int. Assoc. Math. Geol. IAMG 2007*, no. January 2007, pp. 216–219, 2007.
- [55] “Feature Scaling | Standardization Vs Normalization.”
<https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/> (accessed Oct. 03, 2021).
- [56] F. Pedregosa, R. Weiss, and M. Brucher, “Scikit-learn : Machine Learning in Python,” vol. 12, pp. 2825–2830, 2011.
- [57] K. Stapor, “Evaluating and comparing classifiers: Review, some recommendations and limitations,” *Adv. Intell. Syst. Comput.*, vol. 578, no. May 2017, pp. 12–21, 2018, doi: 10.1007/978-3-319-59162-9_2.
- [58] F. J. Valverde-Albacete and C. Peláez-Moreno, “100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox,” *PLoS One*, vol. 9, no. 1, 2014, doi: 10.1371/journal.pone.0084217.

APPENDICES

Appendix A

Sample code

Decision Tree algorithm

```
from sklearn.model_selection import train_test_split
#from sklearn.preprocessing import MinMaxScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

crop_data = df
X = crop_data.drop(columns=['Crop'])
y = crop_data['Crop']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

model = DecisionTreeClassifier()
model.fit(X_train, y_train)
predictions = model.predict(X_test)

score = accuracy_score(y_test, predictions)

print('Decision tree Accuracy on training set: {:.2f}'.format(model.score(X_train, y_train)))
print('Decision tree Accuracy on test set: {:.2f}'.format(model.score(X_test, y_test)))
print('Decision tree Accuracy on predictions :')
score
```

```
Decision tree Accuracy on training set: 1.00
Decision tree Accuracy on test set: 0.99
Decision tree Accuracy on predictions :
0.9911111111111112
```

K-nearest neighbours algorithm

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(X_train_scaled, y_train)
result = knn.score(X_test_scaled, y_test)
print('Accuracy on prediction: {:.2f}'.format(result))
print('KNN Accuracy on training set: {:.2f}'.format(knn.score(X_train, y_train)))
print('KNN Accuracy on test set: {:.2f}'.format(knn.score(X_test, y_test)))
```

```
Accuracy on prediction: 1.00
KNN Accuracy on training set: 0.27
KNN Accuracy on test set: 0.30
```

Data normalization process

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)

scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

X_train_scaled
```

```
array([[0.55789474, 0.46861925, 0.48      , ..., 0.9860114 , 0.01241871,
        0.63636364],
       [0.19473684, 0.28870293, 0.59555556, ..., 0.28939778, 0.09449782,
        0.63636364],
       [0.26842105, 0.12552301, 0.00888889, ..., 0.51060079, 0.06275456,
        0.72727273],
       ...,
       [0.4      , 0.10878661, 0.00444444, ..., 0.27960864, 0.05250136,
        0.63636364],
       [0.55789474, 0.32217573, 0.43111111, ..., 0.85449011, 0.01258225,
        0.90909091],
       [0.4      , 0.18410042, 0.02666667, ..., 0.90296821, 0.10011698,
        0.72727273]])
```

Random Forest algorithm

```
from sklearn.ensemble import RandomForestClassifier

clf = RandomForestClassifier(max_depth=4, n_estimators=100, random_state=42).fit(X_train, y_train)

print('RF Accuracy on training set: {:.2f}'.format(clf.score(X_train, y_train)))
print('RF Accuracy on test set: {:.2f}'.format(clf.score(X_test, y_test)))
```

```
RF Accuracy on training set: 1.00
RF Accuracy on test set: 1.00
```

Gradient boosting algorithm

```
from sklearn.ensemble import GradientBoostingClassifier
grad = GradientBoostingClassifier().fit(X_train, y_train)
print('Gradient Boosting accuracy : {}'.format(grad.score(X_test, y_test)))
```

```
Gradient Boosting accuracy : 0.9911111111111112
```

Support vector machine algorithm

```
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

svc_linear = SVC(kernel = 'linear').fit(X_train_scaled, y_train)
print("Linear Kernel Accuracy: ",svc_linear.score(X_test_scaled,y_test))

svc_rbf = SVC(kernel = 'rbf').fit(X_train_scaled, y_train)
print("Rbf Kernel Accuracy: ", svc_rbf.score(X_test_scaled,y_test))

svc_poly = SVC(kernel = 'poly').fit(X_train_scaled, y_train)
print("Poly Kernel Accuracy: ", svc_poly.score(X_test_scaled,y_test))
```

```
Linear Kernel Accuracy: 0.995
Rbf Kernel Accuracy: 1.0
Poly Kernel Accuracy: 1.0
```

Appendix B

Sample questionnaire

PART A: DEMOGRAPHIC INFORMATION

Q1 Name: _____ (Optional)

Q2 Gender: Male Female other

Q3 Age bracket?

Under 18 18-35 36-45 46-55 56-65 above 66

Q4 Highest Education Level:

None Primary Secondary Post-Secondary

Q5 What is the name of your village/town?

Q6. What is the main economic activity in your village/town?

Q7 How long have you lived in this community?

Less than 5 years 5- 10 years 10-20 years over 20 years

Q8 Do you own a phone or have access to phone?

Yes, I own a phone

No, I do not own a phone

No, I do not own a phone, but I have access to a phone

Q9 If you own a phone, is this a smart phone?

Yes No

Part B: Knowledge on Weather Forecasting

Q10. Do you check weather forecast information?

Yes No

If yes, how often do you check it?

Daily Weekly Monthly Seasonal

Q11. How long have you been using weather forecast information?

Never Less than 1 year 2 to 5 years Over 5 years

Q12. Does the weather and climate information received adequately meet your requirements? (You may select many requirements or none from the requirements given below). Please, tick on a scale 1 –5, with one (1) being the lowest level of adequacy and five (5) being the highest.

	1	2	3	4	5
Disease prediction (for example, malaria prediction)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Farming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Travelling	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Plan the day (like, how to dress, take umbrella or not)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other. Specify please.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

--

PART C: INDIGENOUS KNOWLEDGE BEFORE CULTIVATION

Q1 Which of the following cropping decisions do you make use of indigenous knowledge to reach? (You can tick more than one)

- When/if to plant; for example. *decide not to plant at all based on very delayed onset of the rainy season*
- What to plant; e.g. *to decide to plant sweet potatoes instead if maize based on the anticipated drought*
- How to plant; e.g. decision to practice mixed cropping
- When to harvest; e.g. *if I know there will be frost next week, I can decide to harvest all my crop before*
- Disposal/selling of products; e.g. *when I know that a drought is imminent, I conserve all my products instead of selling them*
- Others, please specify: _____

Q2 Other than indigenous knowledge, do you make use of weather forecasts from the sources below to help you in reaching cropping decisions? If so, do you have confidence in the accuracy of information you get from these options? Please, tick on a scale 1 –5, with one (1) being the lowest level of confidence and five (5) being the highest.

	1	2	3	4	5
Radio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
TV	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Newspaper/Print media	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Website/APP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Local observations. For example, the clouds and behaviour of Animals or/and plants	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other. Specify please.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q3 Using indigenous knowledge, it is simple and reliable to foresee the agricultural season (season with fine weather and good rains) before I cultivate

Strongly disagree disagree Not Sure agree Strongly agree

Q4 Using indigenous knowledge, I can foresee bad season (season with no rains and undesired temperature) before I cultivate

Strongly disagree disagree Not Sure agree Strongly agree

Q5 With indigenous knowledge in place, there is no need for any form of technology that can help to foresee agricultural season

Strongly disagree disagree Not Sure agree Strongly agree

PART D: INDIGENOUS KNOWLEDGE ON CROP MONITORING

Q1 What general crop properties do you observe when evaluating crops other than:

Leaf size, leaf strongness, leaf greenness, stem thickness, stem strongness,

Q2 What soil properties do you observe when evaluating crops other than:

Soil moisture, soil health, soil type

Q3 What environmental properties do you observe when evaluating crops other than:

temperature, wind speed, wind direction, rains, level of crop pests (insects, birds and weed)

Q4 Using indigenous knowledge, it is easy and quick to monitor the development of my crops

Strongly disagree disagree Not Sure agree Strongly agree

Q5 Using indigenous knowledge(observations), I can always spot all factors that may impact my crops such as crop insects and weeds at an early stage

Strongly disagree disagree Not Sure agree Strongly agree

Q6 Using indigenous knowledge, I always experience good crop yield

Strongly disagree disagree Not Sure agree Strongly agree

Q7 There is no need for any form of technology to help me to monitor the development my crops

Strongly disagree disagree Not Sure agree Strongly agree

Q8 How do you protect crops from insect pests and diseases?

Q9 How do you make sure the plant is not water stressed?

Q10 How do you protect your plants from animal attacks?

Q11 How do you protect crops from e.g. hail, frost and floods?

Q12 How can you tell that your plants need fertiliser?

Q13 Is there any indicators you use to foresee crop pests?

Yes No

If yes, what are those indicators?

PART E: AWARENESS OF CLIMATE CHANGE AND BIODIVERSITY DEGRADATION

Q1 In relation to Climate change, please state to what extent you agree with the following statements? Please, tick on a scale 1 –5, (1) for Strongly Disagree (2) for Disagree, (3) for NOT SURE, (4) Agree and (5) Strongly Agree.

	1	2	3	4	5
I have heard of climate change before	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Climate change is real, I have seen some of the effects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Climate change is measurable, Increase in the average temperature of earth's atmosphere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Change in weather condition over an extended period of time is climate change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Climate change is characterized with high temperature	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Climate change comes with rise in sea level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The rate of sunshine is higher now than before	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The weather seems to be hotter nowadays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The weather seems to be colder nowadays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There is increased rate of rainfall	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cases of flooding occur more nowadays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I believe that there is still plenty of time to prepare for climate change problems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think climate change will bring a period of great adversity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am seriously concerned with what problem climate change may bring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think no special preparation is needed for climate change as our forefathers experienced this too	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I believe climate change will pass like other environmental problems, so there is no need to worry	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I am working hard to educate my friends/ family/ neighbours on climate change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I always ask questions on on climate change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I do not think that global climate is changing as our forefathers experienced this too	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q2 In relation to Biodiversity Degradation, please state to what extent you agree with the following statements? Please, tick on a scale 1 –5, (1) for Strongly Disagree (2) for Disagree, (3) for NOT SURE, (4) Agree and (5) Strongly Agree.					
	1	2	3	4	5
I have witnessed biodiversity loss e.g. some trees and animals that I used to see are no longer existing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Human responses to climate change e.g., changes in agriculture, resource conflicts, and migration affect biodiversity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Biodiversity shapes social resilience the effects of climate change	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Genetically modified organisms have Have bot direct and indirect impacts on biodiversity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The different types of protected areas e.g., strict nature reserves, hunting reserves, and national parks, are effective at conserving biodiversity and providing ecosystem services	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Spatial pattern of human settlement e.g., clustered vs. dispersed has the least impact on biodiversity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Agricultural intensification contributes to conserving overall biodiversity by reducing pressure to convert natural ecosystems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Urban nature reserves and other green amenity spaces, such as golf courses, contribute to biodiversity conservation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The weather seems to be colder nowadays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The different strategies devised to integrate scientific knowledge into biodiversity conservation are effective	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	<p>Relationship between economic growth and biodiversity vary across scales, among different types of ecosystems, and with the type of economic activity <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
	<p>Economic subsidies affect biodiversity within the recipient region and elsewhere <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
	<p>Corruption influence the effectiveness of biodiversity conservation <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
	<p>There is a relationship between individuals learning about environmental problems and their conservation attitudes, knowledge, beliefs, and behaviours <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
	<p>Current changes in human patterns of food consumption have effects on biodiversity e.g., shift from bush meat to domestic meat and from fish to plant based protein <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
	<p>Formal recognition of customary rights and traditional institutions should be the basis for biodiversity conservation policy and practices <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>