**CHAPTER ONE**

**INTRODUCTION**

* 1. **Background to the Study**

Livestock is an important agricultural sector that contributes significantly to the economies of countries and livelihoods of several people around the world. Livestock systems also have a large environmental footprint. They use 30% of the global land, 32% of water and emit around 18% of greenhouse gas emissions. The last two decades have seen a proliferation of work on livestock and climate change. The two main threads have been on mitigation of greenhouse gases and on adaptation to climate change. Initially these research agendas were studied independently. Recently, there is growing demand from decision-makers and policy to address them in a more integrated manner as their agenda has moved towards the development of solutions (FAO, 2021).

The study presented an optimized approach to the monitoring of livestock farming via a smart model using fuzzy logic technique and the internet-of-things (IoT). Livestock farming can be defined as the management and breeding of domestic, livestock or farm animals for the purpose of obtaining their meat and products (milk, eggs, leather, etc.) Furthermore, livestock farming is one of the oldest economic activities of man started by early men. It guarantees food supply, hides, skins, bones, milk and other animal products without going to the forest to hunt. Livestock farming includes the breeding of cattle, sheep, pigs, goats, poultry, rabbits, snails, fishes, and honeybees.

Livestock farming comes with challenges such as exploitation by farmers, issues of expansion, economic volatility, input volatility, consumer retailer perception, poor animal health care and so on. The study intends to address the mentioned challenges with the Internet-of-Things (IoT). IoT is a reﬂective system of connecting the real-world objects that are accessed via the network. Sensors are involved for the data collection, in which the objects are having their own IP address and having the ability to access and transmit the resources through the Internet without manual performance.

It is an architectural framework that offers integration and transfers the information between the computational devices. Its application areas are wide spread across smart homes, wearable, automation of vehicles, industrial Internet, smart cities, smart agriculture, smart retail, energy engagement, poultry and farming (Chen et al, 2018)

Fuzzy logic is a branch of science that is extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Several authors and researchers have addressed the issue of smart livestock farming.

For instance, Michael and Gregory (2017), opined that smart farming envisages the harnessing of information and communication technologies as an enabler of more efﬁcient, productive, and proﬁtable farming enterprises. Such technologies do not sufﬁce on their own; rather they must be judiciously combined to deliver meaningful information in near real-time.

* 1. **Statement of the Problem**

Livestock farming comes with challenges such as exploitation by farmers, issues of expansion, economic volatility, input volatility, consumer retailer perception, poor animal health care and so on. The study intends to specifically focus on exploitation of livestock farming by farmers. This specific issue is associated with the concept of factory farming. Factory farming is the main cause of animal suffering and abuse. These silent victims have been converted into machines that generate meat, milk, and eggs. These animals are sentient beings with a desire to live, but are cruelly treated by the farmers that are supposed to manage them. Factory farming operations are designed to produce large volumes of yield for the smallest possible price. It's expensive to farm animals because animals require constant supplies of food, water, and shelter in order for them to grow large enough to be slaughtered, or to produce milk or eggs for human consumption. Factory farming is the result of techniques for keeping animals alive and producing at the lowest costs possible, using cost-saving measures such as smaller cages and extreme confinement. The impacts on the lives of animals in factory farms are significant. Animals bear the burden of the cost-saving measures on factory farms. Their lives are defined by an inability to engage in natural behaviors; being maimed. Todor and Juri (2017), addressed the mentioned problem with an expert system for milk and animal monitoring. The study presented design and development of an expert system for data collection, analysis and decision. This study intends to enhance farm monitoring with the improvement of the existing monitoring system developed by Todor and Juri (2017).

* 1. **Aim and Objectives of the Study**

The aim of this study is to develop a fuzzy-based model for monitoring smart livestock farming. The specific objectives of the study include to:

1. design a monitoring system for livestock farming.
2. optimize the performance of the proposed system with fuzzy logic.
3. implement the new system with python programming and MySQL.
4. evaluate the performance of the proposed system using confusion matrix.
5. compare the existing system of Todor and Juri (2017) with our proposed system.

**1.4 Significance of the Study**

The benefit of the study is to track farm implementation activities, outputs, and effectiveness of farmers’ productivity and efficiency towards farm cattle. Furthermore, the study will be useful in determining farmers’ input to management of farm cattle. In addition, agricultural agencies and researchers with keen interest in the study area will find the study useful.

**1.5 Scope of the Study**

The direction of the study is focused on improved monitoring and health care of the farm cattle by the farmers.

* 1. **Motivation for the Study**

The motivation for the study was triggered by the love for the Internet-of-Things (IoT). The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

* 1. **Definition of Terms**

The following are defined terms that sheds light on the study:

1. **Livestock farming**

Livestock Farming can be defined as the management and breeding of domestic, livestock or farm animals for the purpose of obtaining their meat and products (milk, eggs, leather, etc.)

1. **Internet-of-things**

The Internet of things (IoT) describes the network of physical objects “things” or objects that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the Internet.

1. **Farm monitoring**

Farm platform makes it possible to monitor the farm constantly, measuring data such as smoke, flooding, air temperature and humidity, the temperature of liquids and power cuts originating from cellars, barns, stables, silos, residential buildings and greenhouses

1. **Fuzzy Logic**

Fuzzy logic can be used to describe how information is processed inside human brains. For example, it can be argued that humans do not know the difference between fat and thin. Five people may be fat and not have the same severity of fatness.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 Concept of agriculture**

The term agriculture refers to the usage of technologies like Internet of Things, sensors, location systems, robots and artificial intelligence on your farm. The ultimate goal is increasing the quality and quantity of the crops while optimizing the human labor used. The connection between all these technologies is the internet of things this is a mechanism for connectivity between sensors and machines, resulting in a complex system that manages your farm based on data received. Thanks to this system, farmers can monitor the processes on their farms and take strategic decisions remotely from their tablet, phone or other mobile device without being on the open fields, in their greenhouse, orchard, vineyard, etc.

Farming" is an emerging concept that refers to managing farms using technologies like IoT, robotics, drones and AI to increase the quantity and quality of products while optimizing the human labor required by production. The Internet of Things (IoT) has provided not only a way to better measure and control growth factors, like irrigation and fertilizer, on a farm, it will change how we view agriculture in its entirely.

The goal of agriculture research is to ground a decision making support system for farm management. Smart farming deems it necessary to address the issues of population growth, climate change and labor that has gained a lot of technological attention, from planting and watering of crops to health and harvesting. Figure 2.1 illustrate a typical smart farming model (Lamik et al, 2011)



Figure 2.1: A typical agricultural model

(**Source:** Lamik et al, 2011)

**2.1.1 Livestock farming**

Livestock farming is simply the management and breeding of domestic, livestock or farm animals for the purpose of obtaining their meat and products (milk, eggs, leather, etc.). Figure 2.2 further illustrates livestock farming. It can also be described as the economic activity that involves raising domestic animals for human consumption and obtaining meat, milk, wool, fur, honey, among others. Livestock farming is one of the oldest economic activities of man started by early men. It guarantees food supply, hides, skins, bones, milk and other animal products without going to the forest to hunt. Livestock farming includes the breeding of cattle, sheep, pigs, goats, poultry, rabbits, snails, fishes, and honeybees.

The breeding, maintenance, and slaughter of livestock, known as animal husbandry, is a component of modern agriculture that has been practiced in many cultures since humanity's transition to farming from hunter-gatherer lifestyles. Animal husbandry practices have varied widely across cultures and time periods, and continue to play a major economic and cultural role in numerous communities. Livestock farming practices have largely shifted to intensive animal farming, sometimes referred to as "factory farming"; over 99% of livestock in most countries are now raised in this way. Intensive animal farming increases the yield of the various commercial outputs, but has also led to negative impacts on animal welfare, the environment, and public health. In particular, livestock especially beef, dairy and sheep stocks, have out-sized influence on greenhouse gas emissions from agriculture. Animal-rearing originated during the cultural transition to settled farming communities from hunter-gatherer lifestyles. Animals are domesticated when their breeding and living conditions are controlled by humans. Over time, the collective behavior, lifecycle and physiology of livestock have changed radically. Many modern farm animals are unsuited to life in the wild.



Figure 2.2: Images of livestock farming

**2.2 Deep learning as a tool for improving smart livestock farming**

Deep neural networks also known as deep learning is neural networks with more than one hidden layers. Deep neural networks facilitate the learning of complex function by a machine through representation-learning method with multiple representations obtained by composing simple but non-linear modules that each transforms the representation at one level (starting with raw input data) into a representation at higher level. The representation learning is a set of methods that allows the machine to be fed with or receive raw data and to automatically discover the representation required for detection or classification.

Deep learning solves this central problem in representation learning by introducing representations that are expressed in terms of other simpler representations and enables the computer to build complex concepts out of simpler ones. In statistical machine learning, a major issue is the selection of an appropriate feature space where input instances have desired properties for solving a particular problem. For example, in the context of supervised learning for binary classification, it is often required that the two classes are separable by a hyper-plane. In the case where this property is not directly satisfied in the input space, one is given the possibility to map instances into an intermediate feature space where the classes are linearly separable. This intermediate space can be specified explicitly by hand-coded features, be defined implicitly with a so-called kernel function, or be automatically learned. In both of the first cases, it is the user’s responsibility to design the feature space. This can incur a huge cost in terms of computational time or expert knowledge, especially with highly dimensional input spaces, such as when dealing with images.

In addition, deep learning also consist of multiple layers of nonlinear processing which can be considered as a relevant choice for decisions by machine learning models. Indeed, some highly nonlinear functions can be represented much more compactly in terms of number of parameters with deep architectures than with shallow ones (e.g. Support Vector Machine). For example, it has been proven that the parity function for n-bit inputs can be coded by a feed-forward neural network with O (log n ) hidden layers and O (n ) neurons, while a feed-forward neural network with only one hidden layer needs an exponential number of the same neurons to perform the same task. Moreover, in the case of highly varying functions, learning algorithms entirely based on local generalization are severely impacted by the curse of dimensionality. Deep architectures address this issue with the use of distributed representations and as such may constitute a tractable alternative.

In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. In an image recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face.

Importantly, a deep learning process can learn which features to optimally place in which level on its own. (Of course, this does not completely eliminate the need for hand-tuning; for example, varying numbers of layers and layer sizes can provide different degrees of abstraction.) The word "deep" in "deep learning" refers to the number of layers through which the data is transformed. More precisely, deep learning systems have a substantial credit assignment path (CAP) depth. The CAP is the chain of transformations from input to output. CAPs describe potentially causal connections between input and output. For a feed-forward neural network, the depth of the CAPs is that of the network and is the number of hidden layers plus one (as the output layer is also parameterized). For recurrent neural networks, in which a signal may propagate through a layer more than once, the CAP depth is potentially unlimited. No universally agreed upon threshold of depth divides shallow learning from deep learning, but most researchers agree that deep learning involves CAP depth which has been shown to be a universal approximator in the sense that it can emulate any function.

Beyond that, more layers do not add to the function approximator ability of the network. Deep models (CAP > 2) are able to extract better features than shallow models and hence, extra layers help in learning the features effectively. Deep learning architectures can be constructed with a greedy layer-by-layer method. Deep learning helps to disentangle these abstractions and pick out which features improve performance. For supervised learning tasks, deep learning methods eliminate feature engineering, by translating the data into compact intermediate representations akin to principal components, and derive layered structures that remove redundancy in representation. Deep learning algorithms can be applied to unsupervised learning tasks. This is an important benefit because unlabeled data are more abundant than the labeled data.

**2.3 Fuzzy Logic**

Fuzzy logic is a branch of science that is extended to handle the concept of partial truth, where the truth value may range between completely true and completely false as shown in figure 2.3 and table 2.1. Fuzzy logic may be applied to many fields, including control systems, neural networks and artificial intelligence (AI). Fuzzy logic can be used to describe how information is processed inside human brains. For example, it can be argued that humans do not know the difference between fat and thin. Five people may be fat and not have the same severity of fatness. Or, one person may appear thin, compared to another, while both are actually fat. Using fuzzy logic, you can assign different logic values for fatness, ranging from 0 to 1, according to severity of fatness. Variables between the extremes of zero and one are closer to the concept of probability, which means there is a major correlation between the science of probability and fuzzy logic. However, fuzzy logic refers to intensity of truth, while probability refers to likelihood. Fuzzy logic is an extension of Boolean logic by Lotfi Zadeh in 1965 based on the mathematical theory of fuzzy sets, which is a generalization of the classical set theory. By introducing the notion of degree in the verification of a condition, thus enabling a condition to be in a state other than true or false, fuzzy logic provides a very valuable flexibility for reasoning, which makes it possible to take into account inaccuracies and uncertainties. One advantage of fuzzy logic in order to formalize human reasoning is that the rules are set in

Since its launching in 1978, the journal Fuzzy Sets and Systems has been devoted to the international advancement of the theory and application of fuzzy sets and systems. Fuzzy sets are also the cornerstone of a non-additive uncertainty theory, namely possibility theory, and of a versatile tool for both linguistic and numerical modeling: fuzzy rule-based systems. Numerous works now combine fuzzy concepts with other scientific disciplines as well as modern technologies. In mathematics fuzzy sets have triggered new research topics in connection with category theory, topology, algebra, analysis. Fuzzy sets are also part of a recent trend in the study of generalized measures and integrals, and are combined with statistical methods. Furthermore, fuzzy sets have strong logical underpinnings in the tradition of many-valued logics. Fuzzy set-based techniques are also an important ingredient in the development of information technologies.



Figure 2.3 Illustration of fuzzy logic (**Source:** Mchillson *et al*, 2019)

Table 2.1: Further illustration of fuzzy logic (Source: Mchillson *et al*, 2019)

|  |  |  |  |
| --- | --- | --- | --- |
| If the light is red…. | If my speed is high… | And if the light is close….. | Then I brake hard…. |
| If the light is red…. | If my speed is low… | And if the light is far….. | Then I maintain my speed… |
| If the light is orange.. | If my speed is average | And if the light is far….. | Then I brake gently… |
| If the light is green… | If my speed is low… | And if the light is close….. | Then I accelerate…. |

Fuzzy rule-based modeling has been combined with other techniques such as neural sets and evolutionary computing and applied to systems and control engineering, with applications to robotics, complex process control and supervision. In the field of information systems, fuzzy sets play a role in the development of intelligent and flexible man machine interfaces and the storage of imprecise linguistic information. In artificial intelligence various forms of knowledge representation and automated reasoning frameworks benefit from fuzzy set-based techniques, for instance in interpolative reasoning, non-monotonic reasoning, diagnosis, logic programming, constraint-directed reasoning, etc. Fuzzy expert systems have been devised for fault diagnosis, and also in medical science. In decision and organization sciences, fuzzy sets has had a great impact in preference modeling and multi criteria evaluation, and has helped bringing optimization techniques closer to the user’s needs. Applications can be found in many areas such as management, production research, and finance.

**2.3.1 Analysis of fuzzy logic algorithm**

Fuzzy logic algorithm is a soft computing paradigm built around human thinking and natural occurrences that offers predicates which are present in nature and similar to those either big or small. This theory simulates human thinking as to how a person makes faster decision. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, and also truth values between "completely true" and "completely false". Furthermore, it can be implemented in hardware, software, or a combination of both. In the present competitive scenario the fuzzy logic system are being adopted by the automotive manufacturers for the improvement of quality and reduction of development time and the cost as well. Fuzzy logic was conceived as a better method for sorting and handling data but has proven to be an excellent choice for many control system applications (Singh & Mishra, 2015).

The potentials and ability of fuzzy logic controls and system to imitate and epitomize human knowledge are strongly determined by guesses and error operator. Singhala, Shah and Patel (2014), opined that fuzzy logic constitutes a non-linear mapping of inputs dataset to a scalar output. Furthermore, they also stated that fuzzy logic basically consist of four components known as fuzzifier, rules, inference engine and defuzzifier. In addition, Gursel (2016) argued that the following fuzzy rules should be infused into the knowledge base and database of a system.

a) Fuzzifier: This component is responsible for the translation of the fuzzy crisp into fuzzy values. The fuzzifier is responsible for the fuzzification of crisp which is the method of transforming a crisp object into a fuzzy set, to a grade of membership function for linguistic variables of fuzzy sets

b) Knowledge base (Rules): This includes the knowledge and decision rules captured from expert know-how of the application area managing the relations between the fuzzy input and output. The rule base consists of the IF-THEN conditions based on expert knowledge.

c) Inference Engine: It has the uncertain rational intelligence to get the fuzzy output. Human decision making is simulated to constitute the engine. The processing of the fuzzy set is carried out here according to the rules of the rule base.

d) Defuzzifier: Here the fuzzy output is further translated into a crisp value which is more useful and comprehensible values deployed in real world scenario. The crisp outputs are values constructed by taking into account all parameters in the fuzzy output interval by employing high degree of membership values.

Membership functions are functions that assign a number or value to each element µ(x) of an input space. By implication, it is a function that maps an input value to its membership value. Therefore the membership function for a given value x indicates the degree of its membership to the fuzzy set. There are different forms of membership such as triangular, trapezoidal, Gaussian or singleton. The type of membership function can be context dependent and can be randomly chosen based on user’s experience (Table 2.2).

In addition, a fuzzy set is also an extension of classical set characterized by a membership or characteristics function which gives to individual object a grade of membership that ranges between zero and one. A fuzzy set therefore have more that an ‘either or’ methodology for membership. For example the set considers a set of tall people in the following ways, an individual with the height of 200cm may belong to a set and another of 110 cm does not. Fuzzy logic applications are in use and deployed across industry and research domain as it uses human-like inference rules for making decision especially when the results required are not totally exact which ordinary hard coded computer programs.

Intuitively, it thus seems that the input variables like in this example are approximately appreciated by the brain, such as the degree of verification of a condition in fuzzy logic. Fuzzy logic can be conceptualized as a generalization of classical logic. In fuzzy logic, which is also sometimes called diffuse logic, there are not just two alternatives but a whole continuum of truth values for logical propositions. A proposition can have the truth value 0.4 and a complement value of 0.5. According to the type of negation operator that is used the two truth values must not be necessarily add up to 1. Fuzzy logic has a weak connection to probability theory. Probabilistic methods that deal with imprecise knowledge are formulated in the Bayesian would have difficulty expressing. From temperature regulations in industries to extreme engineering where inaccurate decision making would portend grave danger and business loss, fuzzy rule have come to the rescue.

Table 2.2 Fuzzy logic algorithm

|  |  |  |
| --- | --- | --- |
| **S/No** | **ALGORITHM** | **FUZZY LOGIC COMPONENT INVOLVED / ACTION** |
| 1 | Define linguistic values and terms | Initialization |
| 2 | Construct membership function | Initialization |
| 3 | Construct rule base | Initialization |
| 4 | Convert crisp into fuzzy values using the membership function | Fuzzification |
| 5 | Evaluate the rules in the rule base | Inference |
| 6 | Combine the result of each rule base | Inference |
| 7 | Convert output to non-fuzzy values | Defuzzification |

One common phenomena in data mining is the uncertainty as data mining attempts to extract important information from large volume of data which most of the time are unknown, fuzzy data mining deployed in time series dataset could provide the association rules for evaluating trends and irregularities in the progression of any given time series event. Fuzzy systems also provides the basic mathematical foundation and logic for the mining and capture of the uncertainties associated with human reasoning and cognition making it easy for the creation of knowledge-based information system with near human decision making accuracy with its application across different domains. In addition, Ansari et al, (2007), opines that the simplicity of fuzzy solution has made it adaptable for usage in mining important data also prominent highpoint of fuzzy system relevant to data mining includes the presence of few fuzzy rules in the rule base and few variables used in each rule.

**2.4 The internet of things as a tool for livestock farming**

The Internet of Things (IoT) is the interconnection via the internet of computing devices embedded in everyday objects, enabling them to send and receive data. By the Internet of Things, objects recognize themselves and obtain intelligent behavior by making or enabling related decisions thinks to the fact that they can communicate information about themselves. These objects can access information that has been aggregated by other things, or they can be added to other services. The Internet of Things (IoT) is the interconnection of physical objects such as vehicles, home appliances, and other items embedded with electronic software, sensors, and connectivity which enable these objects to connect and exchange data. Secondly, the general concept of the Internet of Things (IoT) is to effectively manage big data of physical objects on the internet. Internet of Things is a new technology of the Internet accessing. By the Internet of Things, objects have the ability to recognize themselves and obtain intelligence behavior via signaled communication.

**2.4.1 Application areas of IoT**

Internet of things promises many applications in human life, making life easier, safe and smart. There are many applications such as smart cities, homes, transportation, energy and smart environment.

1. **Cities**

Many major cities were supported by smart projects, like Seoul, New York, Tokyo, Shanghai, Singapore, Amsterdam, and Dubai. Smart cities may still be viewed as a city of the future and smart life, and by the innovation rate of creating smart cities today, it will become very feasible to enter the IoT technology in cities development. Smart cities demand requires careful planning in every stage, with support of agreement from governments, citizens to implement the internet of things technology in every aspect. By the IoT, cities can be improved in many levels, by improving infrastructure, enhancing public transportation reducing traffic congestion, and keeping citizens safe, healthy and more engaged in the community. By connection all systems in the cities like transportation system, healthcare system, weather monitoring systems and etc., in addition to support people by the internet in every place to accessing the database of airports, railways, transportation tracking operating under specified protocols, cities will become smarter by means of the internet of things.

1. **Homes and Buildings**

Wi-Fi’s technologies in home automation have been used primarily due to the networked nature of deployed electronics where electronic devices such as TVs, mobile devices, etc. are usually supported by Wi-Fi. Wi-Fi have started becoming part of the home IP network and due the increasing rate of adoption of mobile computing devices like smart phones, tablets, etc. For example, a network to provide online streaming services or network at homes, may provide a means to control all of the device functionality over the network.

At the same time mobile devices ensure that consumers have access to a portable ‘controller’ for the electronics connected to the network. Both types of devices can be used as gateways for IoT applications. Many companies are considering developing platforms that integrate the building automation with entertainment, healthcare monitoring, energy monitoring and wireless sensor monitoring in the home and building environments. By the concept of the internet of things, homes and buildings may operate many devices and objects smartly, of the most interesting application of IoT in smart homes and buildings are smart lighting, smart environment and media, air control and central heating, energy management and security.

Wireless sensor networks (WSNs) with integration to the internet of things technology will provides an intelligent energy management in buildings, in addition to the obvious economic and environmental gains. Internet together with energy management systems also offers an opportunity to access a buildings’ energy information and control systems from a laptop or a Smartphone placed anywhere in the world. The future Internet of Things will provide an intelligent building management system which can be considered as a part of a much larger information system used by facilities managers in buildings to manage energy use and energy procurement and to maintain building systems.

1. **Energy, health and grid**

A smart grid is related to the information and control and developed to have a smart energy management. A smart grid that integrate the information and communications technologies (ICTs) to the electricity network will enable a real time, two way communication between suppliers and consumers, creating more dynamic interaction on energy flow, which will help deliver electricity more efficiently and sustainably. The key elements of information and communications technologies will include sensing and monitoring technologies for power flows; digital communications infrastructure to transmit data across the grid; smart meters with in home display to inform energy usage.

A close attention that required to hospitalized patients whose physiological status should be monitored continuously can be constantly done by using IoT monitoring technologies. Many people around the worlds are suffering from bad health because they do not have ready access to effective health monitoring and may be suspected to be as critical situation patients. But with small, powerful wireless solutions connected through the IoT is now making it possible for monitoring to come to these patients. These solutions can be used to securely capture patient’s health data from a variety of sensors, apply complex algorithms to analyze the data and then share it through wireless connectivity with medical professionals who can make appropriate health recommendations. The development in transportation is one of the factors to indicate the wellbeing of the country. A road condition monitoring and alert application is one of the most important areas of IoT transportation application. The main idea of the concept of smart transportation and mobility is to apply the principles of crowd sourcing and participatory sensing. The process began with user identifies the route he/she wishes and marked some points as pothole in the smart phone's application The smart transportation deals with three main conceptions (Ibikunle, 2013)

1. **Factory**

Smart factory added a new value in manufacturing revolution by integrating artificial intelligence, machine learning, and automation of knowledge work and communication with the manufacturing process. The smart factory will fundamentally change how products are invented, manufactured and shipped. Industries and manufacturing revolution became one of the most developed technologies nowadays, the growth of the industry evolution taken many generations. The first generation related to the mechanical machines in addition to water and stream power. The second industry generation deal with mass production, assembly lines and electricity.

In the end of the last century, industries operated under control of computers and automation which recognized as third generation of industries. Environment plays a major effect in human life. People, even animals, birds, fishes and plants may be affected in unhealthy environment. The environment needs smart ways and new technologies for monitoring and management. Monitoring the environment is important in order to assess the current condition of the environment, to take correct life decision according to collected data from monitoring systems, and management is needed to have an efficient resources consuming and use in addition to decrease the factories and vehicles wastes (Adeshina et al, 2016)

**2.4.2 Major internet-of-things challenges**

Major challenges encountered during embedded computing via the internet of things application involve the following:

**i) Security**

IoT has already turned into a serious security concern that has drawn the attention of prominent technological firms and government agencies across the world. The hacking of baby monitors, smart fridges, thermostats, drug infusion pumps, cameras and even the radio in your car are signifying a security nightmare being caused by the future of IoT. So many new nodes being added to networks and the internet will provide malicious actors with innumerable attack vectors and possibilities to carry out their evil deeds, especially since a considerable number of them suffer from security holes.

**ii) Connectivity**

Connecting so many devices will be one of the biggest challenges of the future of IoT, and it will defy the very structure of current communication models and the underlying technologies. At present we rely on the centralized, server/client paradigm to authenticate, authorize and connect different nodes in a network.

**iii) Compatibility and longevity**

IoT is growing in many different directions, with many different technologies competing to become the standard. This will cause difficulties and require the deployment of extra hardware and software when connecting devices.

**iv) Standard**

Technology standards which include network protocols, communication protocols, and data-aggregation standards, are the sum of all activities of handling, processing and storing the data collected from the sensors. This aggregation increases the value of data by increasing, the scale, scope, and frequency of data available for analysis.

**2.5 Review of related works**

Todor and Juri (2017), looked at Expert System for milk and animal monitoring. The study presented the design and development of an expert system for data collection, analysis and decision. Furthermore, the authors also opined that Expert systems (ES) are characterized with high performance, high responsiveness and reliability combined with ease of use and ease of understanding, which makes developing of such system challenging and involves different type of specialists.

Abraham et al, (2019), looked at a Machine Learning Approach for Smart Livestock IoT. The authors analyzed various existing supervised and unsupervised machine learning techniques applied in agricultural domain and compare one technique with another with respects to accuracy and a confusion matrix is plotted for each.

Andzio and Xiong (2019), looked at Livestock Monitoring System using Smart and Innovative Livestock Farming. The authors developed an optimal crop irrigation/agriculture system based on a network of wireless sensors. The work intended to strategize and maintain a control organism utilizing crop sensor that senses with data managing with a web application and a smartphone.

Reeve (2015), looked at Climate Challenges affecting Livestock farming in India. The authors opined that Global climate change is emerging as a huge challenge for human and animal survival on Earth. With increasing population, industrialization and other development related activities, tremendous pressure is exerted for improved production and productivity from agriculture and allied sectors.

Veronica & Francisco (2019) looked at Smart Livestock Farming towards Agriculture 5.0. The study reviewed the current status of advanced farm management systems by revisiting each crucial step, from data acquisition in crop ﬁelds to variable rate applications, so that growers can make optimized decisions to save money while protecting the environment and transforming how food will be produced to sustainably match the forthcoming population growth.

Fontios et al, (2019), looked at a Review of Machine Learning and IoT in Smart Transportation. The authors carried out self-contained review of ML (Machine Learning) techniques and IoT applications in Intelligent Transportation Systems (ITS) and obtain a clear view of the trends in the aforementioned fields and spot possible coverage needs. From the reviewed articles it becomes profound that there is a possible lack of ML coverage for the Smart Lighting Systems and Smart Parking applications.

Park et al, (2019), proposed Clot-Net: A Scalable Cognitive IoT based Smart City Network and Architecture. The authors proposed a CIoT-based smart city network (CIoT-Net) architecture which describes how data gathered from smart city applications can be analyzed using cognitive computing and handle the scalability and flexibility problems. Furthermore, the authors also discussed various technologies such as AI and big data analysis to implement the proposed architecture.

Fatima et al, (2019), looked at Machine Learning in IoT Security: Current Solutions and Future Challenges. The authors systematically reviewed the security requirements, attack vectors, and the current security solutions for the IoT networks. Furthermore, they shed light on the gaps in security solutions that call for ML (Machine Learning) and DL (Deep Learning) approaches.

Jamie et al, (2017), looked at Healthcare in the Smart Home: a study of past, present and future. The study looked at the history of Smart Home Healthcare, current research areas, and potential areas of future investigation. The authors did a good job. However, the reviewed history on healthcare was not practically implemented to a model.

Marijke et al, (2018), looked at activities of daily living in older community dwelling persons. The authors analyzed Electronic databases (Medline, EMBASE, AMED, Psycinfo, CINAHL), using MeSH terms and relevant keywords. Studies, published in English, were included if they evaluated one or more psychometric properties of ADL instruments in community-dwelling older persons aged 60 years and older. The authors did a good job. However, they failed to utilize a biometrical approach to the integration of health-related electronic databases.

Sachin & Durga, 2015 looked a data mining framework to analyze road accident data. The study proposed a framework that used K-modes clustering technique as a preliminary task for segmentation of 11,574 road accidents on road network of Dehradun (India) between 2009 and 2014 (both included). Next, association rule mining are used to identify the various circumstances that are associated with the occurrence of an accident for both the entire data set (EDS) and the clusters identified by K-modes clustering algorithm. The findings of cluster based analysis and entire data set analysis are then compared. The results reveal that the combination of k mode clustering and association rule mining is very inspiring as it produces important information that would remain hidden if no segmentation has been performed prior to generate association rules. The authors did a good job. However, their model was produced inaccuracies and latencies during their mining and clustering process for road accidents prediction. In addition, the mentioned limitation in their developed frame work was due to the absence of a Case-based recommender technique.

Ranjan et al, (2015), compared the various techniques adopted for use in carrying out the process of classification. Such models like k-nearest neighbour (KNN), neural networks (NN), bayes method (BM), support vector machine (SVM), decision trees (DT) were used. When the comparison was made, it was identified that the simplest method that needs to be adopted for use is k-nearest neighbor algorithm. This method is not good for high dimensions documents. Neural network and SVMs performed better than others as they are applied on multi dimensions. Sample size that was larger were needed so that maximum accuracy will be achieved while using neural network and SVMs. Data set that was relatively lesser and storage space that is little are needed by the naive Bayes. KNN and NN do not tolerate noise, classification that is association based and decisions resist noise. Comparing other classifiers, SVM had high learning speed; irrelevant features tolerance is high, among others. Yet recommending any technique to be superior to others is very difficult one. This is because the choice of technique for modeling depends on what the organization wants and the available data. More research is recommended so as to find out the relationship that exists among these models.

Shen *et al,* (2016), proposed an approach through which document can be represented using sequences that are extracted automatically through the use of n-multigram models. Two text classifiers were on the proposed presentation of document. While one was implemented through the use of n-multigram model so that the document can be classified as as sequences that depend on representation are generated the same period of time of time, the other was achieved as the n-multigram models is combined with n-gram models. There are three stages in which this algorithm operates in. The first stage was a period of training on the documents using the n-multigram models for each of the category. There is the splitting of the documents for training into sequences based on the n-multigram models. The stage that follows the above has the text classifiers that are dependent on n-gram model being trained on the steps that the first stage came up with. Stage three does the classification of the test documents. In each of the test document, the models (n-multigram model and n-gram model) each appearing in pair, and obtained from each category will be carried out on the document for the test. Sharing into sequences, of the document takes place first to the n-multigram model from a particular model. Assigning the test document was done to the category whose probability is largest. A series of series of experiment was conducted on RCV1’s subset. The result indicated how well the proposed text classification algorithms performed. Even though the size of the model proposed algorithm that depended on the n-multigram models is smaller than the n-gram models dependent classifier, similar or classification that is better in performance can be achieved. The results indicated as well how the algorithm that was proposed, i.e. that which is based on the n-multigram models combination and n-grams models brought an improvement to the values of micro-F1 and macro-F1 ranging to 92.6% from 89.5% and to 91.1% from 87.2% respectively are classifiers which depend on n-gram models are compared. 3-fold cross validation methodology was used to carry out the experiment. To determine the highly significant improvement which was α= 0.0003, the t-test was used. Suggestion was made on the use of other classification algorithms like SVM and KNN depending on the proposed document representation approaches to determine its validity, and also to carry out experiments using some other data sets so that its adaptability will be verified, while carrying out further researches.

Roa *et al,* (2016), proposed an intelligent text data classification system (ITDCS). The system was designed on the inspiration of Biology using Genetic approach. It focused on how the classifier can be modelled from the data for training considering the genetic algorithm and the appropriate components that are found in Biology. The reason for this is to enable the accurate acquisition of computational intelligence. From the on-set ITDCS is determined to get structured data prepared through the use of unstructured data that is in huge volume using procedural steps in conjunction with filter methods. K-nearest neighbor (KNN) based classification was used to classify the data text into classes labeling so that the best features got by the genetic algorithm will be selected. In doing this, it specifically added to the classifier the power obtained from intelligence joining parts obtained from biology. Such parts are: strategy for encoding, fitness function, and genetic algorithm operators. The putting together of all the component of biology in genetic algorithm in Intelligent data classification system (ITDCS) made the accuracy and efficiency to improve significantly and brought the rate of misclassification to a great reduction while the text data is being classified. The challenge that was open was the application of the techniques for classification and clustering of text data. Therefore a suggestion was made for future research. Improving the performance of genetic repetition, genetic algorithm through the use of parallel processing was also mentioned for future research.

Al-Khurajji & Sameh, (2016), proposed an approach for use in Arabic text classification using a classifier known as Kernel Naive Bayes (KNB). The following techniques for pre-processing were used. They are: Arabic words light Stemmer, Stop word removal, and word tokenization. TF-IDF technique was applied as well for Arabic words feature extraction so that they will be converted to vector space for normalized classification. Dataset was made up of Arabic topic mining corpus. It contains 1897 documents from topics which emanated from 3 areas namely sports (633 documents), culture (639 documents), and Economic (625 documents). The corpus has 2478 words that are unique. Waikato Environment for Knowledge Analysis (WEKA), a machine learning and data mining tool developed in New Zealand at the University of Waikato was used to conduct the experiment. The results showed that the proposed classifier got good results considering the accuracy and time which is contrary to other classifiers used in the previous studies. It is hereby concluded that Arabic text classification of electronic document through the use of the KNB that was proposed indicated better performance than others.

Hmeidi *et al,* (2016), in a set of experiments compared the performance of five best algorithms that are known in text categorization. That study was aimed at making sure that Arabic language text categorization problem was solved. The experiment commenced with the selection of five categories from Arabic dataset that is popular. Three different versions which emanated from the dataset that was considered to be the source through other either root-based stemmed or light10 stemmed were used. Finally, on the dataset were conducted several experiments through WEKA, and using classifiers that are popular like Decision Table classifiers, Decision Tree, KNN, NB, and SVM. The accuracy and scalability of two data mining and machine learning tools (WEKA and RapidMiner) were used so that the prospects and constraints of their use for the categorization of Arabic text will be determined. From the result, SVM performed more than others. Moreso, light10 stemmer yielded better results than the root-based stemmer, considering accuracy. They suggested the use of other classifiers as well as techniques for feature reduction in future researches for further comparison so that deeper insights will be obtained. And finally, they recommend that resourced of Arabic text categorization should adopt the use of RapidMiner because it is scalable and effective.

Adel *et al,* (2016), evaluated K-NN, C4.5 and Rocchio classifier in a comparative study. Arabic data set was where the evaluation was carried out on. The dataset that were used in carrying the experiment was taken from the web of Aljazeera news, Al-hayat, and Saudi Press Agency. The dataset was made up of 1400 Arabic documents from different categories. These dataset were shared into parts. The other being 480 documents with a percentage of 34 was used in carrying a testing. Evaluation measures such as F1, Precision, and recall were used by the authors. Another name for precision is positive predictive value. Precision in information retrieval concerns itself with number of true positive meaning the amount of items that are labeled correctly to have found itself in the positive class which is divided using total number of the labeled element found in the positive class. While Recall is said to be the true positives’ number that is divided through the use of total number of elements which truly is of the positive class. The test’s accuracy is determined using F1. The categories were: Computer, Education, Medicine, Religion, Economics, Law, Politics, and Sports. Results of the experiments showed that the recall and average precision through the use of Rocchio classifier out performed C4.5 on the dataset that were selected. K as a value is significant on K-NN algorithm. Small value of K will bring about low performance of precision and recall. And conversely, if the value of K is high, it means there is going to be a computation that is very high. It summarily shows from the result that Rocchio and K-NN can perform very creditably on Arabic data set. It also said that application of the algorithm on classification using Arabic dataset requires more or further investigation (research).

Thabtah *et al,* (2016), carried out a comparison between Vector Space Models (VSMs) of different variations and term weighting approaches through the use of KNN algorithm. The reason for that comparison was to carry out an F1 measure of evaluation. The data adopted for use in that experiment were datasets obtained from Arabic online newspapers which include Al-Dostor, Al-haran, Al-hayat, Al-Nahar, and Al-Jazeera. Arabic text being translated from English as Arabic language is a derivational one and highly inflectional. This makes it difficult for monophonical analysis to take place. Again, representation of vowels is done through diacritics in Arabic script. This usually leaves out of the text and uses capitalization for proper noun which in the text creates ambiguity. Three Text classification/categorization techniques which depend on vector model similarity (Dice, Jacard, and Cosine) were compared in terms of F1 measure. The same strategy was used while the use methods for classification of text is incoming, KNN comparison was made through the use of different term weighting such as WIDF, Log (1+tf), ITF, and KNN. The categories were Science, Health, Art, Politics, Economy, and Agriculture. The results obtained from three categorizers (Dice, Jaccard, and Cosine) against six Arabic datasets, as in each of them, they considered 70% of the documents arbitrarily for the purpose of training where testing was carried out using 30%. The K parameter was set to 11, in KNN algorithm. It was discovered that among others that consistency existed on Dice- based on TF.IDF and Jaccard based on TF. IDF algorithm whereby both of them performed more than cosine based TF.IDF, etc. Also discovered were similarities existing between Jaccard methods and that of Dice, as similar F1 results were produced by both of them for weighting that has the same term. There was a proposal that in future a new multi-label classification approach that is dependent association rule for the Text Classification/Categorization (TC) problem will be carried out and also an intension to build Arabic language whose datasets for text classification/categorization (TC) will be larger.

Roiss & Nazlia, (2016), investigated the use of K-nearest neighbour (KNN) classifier with the combination of an Inew, Cosine, dice, and Jaccard similarities. The reason for that experiment was to enhance Arabic Text Classification. The dataset that were used in that study were made up of 3172 documents. These documents were found into categories of four namely: Politics, Sport, Art, and Economic. These dataset which were shared in 1732 training and 1440 test documents set respectively were obtained from the website. Three types of data were used to conduct four experiments. These experiments measured the performance of Inew similarity against the Cosine, dice, and jaccard similarities. Three phases were considered. They were: Unstemmed data for (Experiment 1), Stemmed data for (Experiment 2), feature selection of four methods (Chi-Square Statistic (CHI), Mutual Information (MI), GSS Coefficient (GSS), and Odds Ratio (OR) Stemmed data (Experiment 3) and the fourth experiment (Experiment 4), that has to do with time measurement considering the last experiments. Datasets having Bag-Of-Word (BOW) was used to represent each of these phase through the use of simple words as N-Gram and features through the use of sequence characters (N-Gram level character) with the n length, and Tri-Gram (3-Gram) were used. The performance of the K-nearest neighbor classifier having Inew, Cosine, dice, and jaccard similarities were investigated, looking at the unstemmed data, stemmed data and each method of feature selection (CHI, MI, GSS, and OR); through the selection of variable number in each feature set (BOW and 3-Gram) of the top most frequent terms. A lot of experiments were run on the K-nearest neighbor classifier, through the use of Inew, cosine, dice, and jaccard similarities. Inew classifier was majorly found to be the best, as it needed less time comparing cosine, dice, and jaccard similarities.

Ogada *et al,* (2016), conducted an experiment for the purposes of demonstrating how their simple modification will make an improvement on the performance of Naive Bayes for classification to be significantly possible. These experiments were carried out using hate speech data in Kenyan language but translated into English. The dataset that were used in that study were texts that were got from social media online and some text data which researchers generated. In order to select data for training and testing, purposive sampling method was adopted. All the experiments in that study were carried out using 63 sentiments of corpus. To carry out an evaluation of that experiment, 53 sentiments were used for training whereas others were for classification. WEKA implementation of Support Vector Machines (SVM), K-nearest neighbor (KNN) and Naives Bayes (NB) classifiers were carried out. They were: word based n-gram features and word based features. Initially the experiments were carried out through the use of word based features and afterwards word based n-gram features were then conducted. Precision, F-Measure, Accuracy, and Recall were used in the evaluation of the three classifiers namely: Support Vector Machines (SVM), K-Nearest Neighbor, and Naive Bayes (NB). Looking at the accuracy results got through the use of training data on the classifiers as word based features are used over n-grams range, using uni-grams (1-gram), K-nearest neighbor had the highest accuracy with a 98.1% value, next is SVM with 96.2% value while NB has 90.6% For bi-grams (2-grams), the three classifiers gave equal accuracy with 98.1% value. KNN never showed an atom of improvement. Using tri-grams (3-grams), SVM and KNN obtained same accuracy as they were in bi-grams with 98.1% while there was a reduction by almost 6% for the NB. From tri-grams, all the classifiers indicated that the accuracy values were reduced. In the analysis on the efficiency of n-grams indicated that 2-grams (bi-grams) had performance which was better for text categorization for NB. K-nearest neighbor obtained the same value of 98.1% accuracy for 1-gram (uni-gram), 2-gram (bi-gram), and 3-gram (tri-gram) when using training data. Whereas SVM had a value of 98.1% as well, this was also high for 2-gram (bi-gram) and 3-gram (tri-gram). These results indicated that n-grams had improved performance using machine learning algorithms.

Li & Wu (2016), studied online forums hotspot and forecast using sentiment analysis and text mining approaches. First of all, to inspect the sentiment polarity for each piece of text, an algorithm was created. Afterwards to develop unsupervised text mining approach the algorithm was joined with k-means clustering and support vector machine (SVM). Described text mining approach had been used to group forums into various clusters, whose center represent a hotspot forum within the current time span. The datasets had been taken from SINA sports forum. Experimental results showed that SVM forecasting gets high consistent results with k-means clustering. The top 10 hotspot forums given by SVM forecasting resemble 80% of k-means clustering results. Both SVM and k-means achieved the same results for the top 4 hotspot forums of the year. In this paper they had created an algorithm that automatically analyzes the sentiment polarity of a text, with the help of which text values were obtained. Influential power of text was represented by absolute value and sentiment polarity by the sign of text. Previously created algorithm was then combined with k-means clustering and SVM classification to integrated approach for online sports forums cluster analysis. Unsupervised algorithm had been applied to group the forums into various clusters, whose center represent hotspot forum with the current time span. In addition to clustering the forums based on data from the current time window, forecasting for the next window was also done by them. Proof for existence of correlations between post text sentiment and hotspot distribution was given by empirical studies. Results showed that both SVM and k-means produce consistent natural groupings. Several companies could be benefited from these hotspot predicting approaches in different ways. These companies could also combine results for market basket analysis to yield comprehensive decision support information. A firm in financial sector or the financial department of a giant company might get profit from such a sentimental and text mining process. In financial market, right before a security market opens and trading begins, analysts people on sales and trading desks usually try to get an overall fix on market sentiment and for particular investments. First, algorithm design could be improved to yield a more accurate calculation of sentiment. Even for supervised learning, algorithms other than SVM, or variations of SVM, could be joined as well. Secondly, they had incorporated topic extraction. Third, a practical system, in the form of a website portal, was desired as their major future work.

Tripathy *et al,* (2016), represented that the reviews and blog datasets obtained from the social networking sites were unsystematic and need classification for meaningful information. They could be classified as positive, negative and neutral with the help of supervised machine learning methods. In that study, for classification of sentiments, they introduced four different machine learning algorithms i.e. Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD) and Support Vector Machine (SVM) based on precision, recall, F-measure, and accuracy. That study helped in classifying the movie reviews using supervised machine learning algorithms which were further applied on IMDB dataset using n-gram approach. They concluded that in n-gram approach as the value of n increases, the classification accuracy decreases. It was also concluded that combination of TF-IDF and count vectorizer techniques helps in obtaining better accuracy. On further studying they also came across some limitations as small size of twitter comments, reviews or comments including punctuation symbols and words like “greatttt, fineee” as they don't have proper meaning. So, new list of words was prepared for classification after removing the stop words to select the best feature. For better accuracy hybrid machine learning techniques were also considered.

Wahbeh & Al-kabi (2016), compared the performances of three classification techniques that are popularly known such as C4.5 classifier, Naive Bayes (NB) classifier, and Support Vector Machine (SVM) classifier. A set of Arabic text document was used. The aim of that text classification was to assign automatically the text to a category that was predefined depending on features that were found in linguistic, as well as content. Data was obtained from three sources such as Kooora website, news-all website, and Saheeh Al-Bukhari book and others (websites). Four categories of document sets were considered in that study. They were: Prophet Mohammed which says (Al-Hadeeth Al-Shareef), politics, economics, and sports. Each of the categories had 250 text documents, giving a total of 1000 corpus in terms of size. In the study, text document went through a preprocessing stage such as stop words removal, some characters’ normalization, removal of text which are non-Arabic, and symbols as well through the use of simple C# software program that was developed by second person in that study. The documents were changed to a suitable file format so that their usage for the running of the classification techniques mentioned above will be made possible. The Waikato Environment Knowledge Analysis (WEKA) toolkit was used to actualize this. The results were obtained for each of the algorithm so that the accuracy of each of the classifier will be measured. The accuracy measure was used in the performance of the comparison between the classifier so that the best among them will be determined.

Gupta & Rani (2016), proposed an improvement in KNN classifier. The main aim of that proposed work was to come up with an improvement in the existing algorithm. The experiments were performed through the use of mini-news group data set obtained from the Archive of UCI (University of California, Irvine) KDD, a repository of large volume of data found online, and covers a wider and different types of data, whose analysis task and areas of application are considered. This mini-news group has 20 groups of 100 document in each of them were used. From the results obtained from that study, it was shown that for accuracy, KNN had 0.77 while iKNN was 0.91; for precision, KNN was 0.67, while iKNN was 0.86; in Recall, KNN had 0.66 while iKNN had 0.86; for F-Measure, KNN was 0.65 and iKNN was 0.85, and finally, for Time, KNN has 0.015, while iKNN was 0.005. It was concluded that from the results obtained, the new algorithm was more efficient and accurate as well than the algorithm that was ready existing.

Abikoye *et al,* (2016), proposed binary text classification using an ensemble of Naive Bayes and Support Vector Machines. The main objective of that study was to combine classifiers together so that better result will be achieved in the classification. The datasets for that experiment were obtained from UCI (University of California, Irvine) repository. It carries 3000 texts instances tagged either as positive and/or negative Sentiment. Those texts were obtained from restaurants review, product, and movies. While the Sentences were obtained from three websites which include the following: yelp.com, amazon.com, and imdb.com. In each of these websites, 500 negative Sentences were respectively contributed. While the data were being collated, it made sure that the positive and/or negative statements were clearly stated in order to avoid statements that are in neutral form. This was termed as Sentiment labeled Instances. Another one was the Short Message Service span collection obtained from public set of SMS labeled messages that were gotten from mobile phone span research. That dataset has SMS messages classified to be either spam and/or ham. Also, that data has up 5574 Sentences. 4827 of them were ham messages while 747 fell under the category of spam messages. Preprocessing was carried out so that the documents will be normalized, tokenized and changed to lowercase format.

That experiment adopted the use of ensemble method. Stacking and Linear regression were used in the combination of two algorithms. Bag of words model was used in the document representation. Whereas feature selection technique adopted was the Term frequency and Inverse document frequency (TF-IDF) chosen due to the fact that it is efficient in the selection of important words effectively thereby making classification to be possible. Waikato Environment Knowledge Analysis (WEKA) toolkit was the data mining software that was used to carry out that research. The process of classification was preprocessing of the data, feature selection, application of individual algorithms for classification, combination of algorithms for classification, and evaluation of results. The result showed that combining MBN (Multinomial Naive Bayes) and SVM (Support Vector Machine) gave an accuracy that was higher, having confidence more in classification than the use of individual algorithms while carrying out tasks on binary text classification.

Kadiru *et al,* (2016), conducted an experiment on Albanian Text Classification using Bag of words model and word analogies. The objective of that study was text classification of Albanian news articles using two approaches. In the methodology, five different Albanian news portals were the sources of where news were extracted, the articles were taken, including the titles they have, the links in which they can be obtained, their sources, summaries as well as category. The news was one-month duration, having news portals whose foundations could be traced to Albania, and Kosovo. Some of this news was taken through the use of RSS feeds which were taken from the sources. Seven categories of documents were used namely: latest-998 articles, economy-60, sports-98, showbiz-62, technology-48, culture-69, and world-95. In the initial approach, the words that were collected were taken as components that were independent. Vector that conform in the vector’s space were shared to each of these words. Nine classifiers namely: multinomial, Linear SVC, Neighbor, Bernoulli, Nearest Centroid, SGD, Perceptron, Ridge, and Passive Aggressive, were taken from the Scikit-Learn Package, were used. Also, training of the classifiers using some of the articles of the news, as well as testing of their accuracy with the other articles left were carried out.

The second approach looked at the text classification words that pose some treats according to their syntactic and semantic similarities of words, assuming there is a formation of word through the use of characters arising from n-gram. Considering the above, these researchers adopted the use of fastText, a classifier that is hierarchical in nature which put into consideration word that is in local order, and information of sub-word. Each classifier was measured separately so that their accuracy will be determined. The training and time for testing were also analyzed. The results indicated that the bag of word model performed more than fastText while the testing stage of the process of classification was going on for dataset of text that was not large. As multi-label text classification was going on, FastText performed better. It was concluded that news articles can be used to test the classification of Albanian text algorithms, as the best result was obtained using a bag of word model having 94% accuracy.

Naji *et al,* (2016), carried out an experiment on Text classification for Arabic words using BPSO/REP – Tree. The objective of that study was for the provision of a new system that will be used for classification of text which depends on BPSO/REP-Tree hybrid. The experiment was carried out on Arabic data set taken from BBC-Arabic website. There were 110 text documents in the dataset. Seven categories were considered, namely: Middle East news, Health, Sport, Computer, Technology, Communities, and Varieties. REP – tree combined with 66% cross validation were for specific training. This was done for the purpose of text classification. The classification process was carried out using Waikato Environment for Knowledge Analysis (WEKA) tool. Three factors of precision, Recall, F-measure were used in the evaluation to determine the performances of the classification process. The result obtained from the process of classification using REP-Tree in the WEKA tool indicated that varieties class had the overall best performance with precision having 99.5, recall-96.7, while F-measure – 90.2 and not being the best but performed better with other factors thereby recording an average that is highest. Technology recorded the next best performance with precision – 99, recall – 92.4 whereas F-measure – 96.8. Middle East news came third with precision – 94.9, recall – 96.2, and F-measure had 92.3. Communications were ranked the fourth with precision -94.3, recall – 95.1, and F-measure – 93.7. Two classes, Sport and Health recorded the worst with precision – 93.1, recall – 91.0, F-measure 95.6; and precision – 78.4, F-measure – 77.3 respectively.

The classification process was carried out on the previous dataset using another classifier, J48 tree. The reason for this was to make comparisons in their performances. From the comparison made, it was indicated in that result that Communications class performed poorly with precision recording 68.3, while recall had 76.1. It was concluded that the proposed REP-Tree be used for Arabic Text Classification as the results that presented showed that it (the classifier) has a high accuracy.

Duwairi (2016), carried out an experiment on Arabic text categorization. The main objective of the study was to compare the performance of three classifiers namely: naive Bayes (NB), K-nearest neighbors (Knn), and distance – based classifiers. In order to carry out the assessment of the accuracy of the classifiers that were being proposed, Arabic text corpus was taken from magazines that were online and newspapers. 1000 documents of different length and styles of writing were obtained. These documents were of 10 pre-defined categories, with each category carrying 100 documents. The pre-defined categories set were: Sports, Internet, animals, religion, medicine, economic, art, plant, technology, and politics. The documents that were collected were categorized manually by two individuals. Every document was given to one category only. Any document identified to have appeared more than others, was given to others having the tendency of being in a maximum likelihood of human categorizer’s judgment. 50 documents were specified randomly for every category. They were used to train while the other 50 were used to test. The expression of accuracy of the classifier was done based on precision, recall, error rate, and fallout. The results indicated that Naive Bayes classifier performed more than K-nearest neighbor and distance – based classifier.

Rajeswari *et al,* (2016), proposed the use of two unique stage feature selection method that would be used for text categorization. They used Information gain (IG), Principle component analysis (PCA), and genetic algorithm (GA). Two stage feature selection and extraction were used so that it will be possible for easy reduction of high dimensionality of a feature space made up of terms of large range, removal of features that are redundant and irrelevant from the space of the feature. As these were being carried out, a decrease was recorded in the computational complexity of the machine learning algorithms that were used for the text categorization, thereby leading to performances increase. Every term in the first stage of the text is categorized according to how important they are in the classification following a decreasing order through the use of Information gain method. Due to this fact, high importance terms are given to the initial categories and terms whose importance were less, were given to the categories that follow. A reduction in the dimension takes place in accordance with the Information Gain (IG) strategies. Terms that were not of importance were ruled out or ignored as the categorization process was going. The terms having highest importance went through feature selection and extraction method. Because of this, a reduction occurred in the computational time, and that of the complexity. To analyze the dimensionality reduction, datasets assortment that could be used as suggested by the experimenters were the Classic4 data set. However, for that experiment, KNN (K-nearest neighbour) and C4.5 decision tree classifiers were used on the dataset that was selected. The results indicated the reduction found in the computational time and complexity of the categorization.

Al-Salemi & Ab Aziz (2016), investigated Bayesian Learning models so that there will be an enhancement in Arabic Automated Text Classification (ATC). In order to carry out that study, dataset that was used were obtained from Arabic news made up of 3172 documents. It covered the following categories: Sports, Politics, Economy, and Arts. For training purpose, there were 1732 documents divided from dataset. While for testing, 1440 documents were used. Pre-processing of the texts that were plain was the first step in the categorization process. It was made up of the following: tokenization, normalization, removal of stop-words, as well as stemming. In order to represent text character-level, n-gram having 3, 4, and 5 length were used, stemming-words inclusive. At the end of the representation of the text, four features sets that were different, were extracted, one for the representation of each methods. Feature selection (FS) methods for the reduction of features dimension were employed. The following were the FS methods that were employed in that study: CHI, MI, OR, and GSS. The following Bayesian Learning Models were built and trained: Naive Bayes (NB), Multivariate Bernoulli Naive Bayes (MBNB), and Multinomial Naive Bayes (MNB). In that study, there were three methods of feature representation, four techniques of feature selection, as well as three classifiers. For each classification, 12 experiments were carried out, giving a total of 36 experiments in the classifier. The performance of each classifier was determined on the test set through the use of different number of frequency terms found in each of the feature set that were at the top most. 200, 400, 600, 800, 1000, and 1200 features were the numbers found at the top selected features. From the result it was indicated that MBNB performed more than MNB and NB, by achieving a high accuracy. When there were larger number of feature, MNB can perform more than NB and MBNB. The 3-gram representation resulted to poor performance.

Al- Harbi *et al,* (2016), carried out experiments on document classification on seven different Arabic corpora. They used statistical methodology to carry out the experiment. The goal of the experiment was the evaluation of two classification algorithms namely: SVM (Support Vector Machine) and C5.0 on Arabic text classification through the use of seven Arabic corpora. Data mining software, two in number namely: RapidMiner and Clementine were used to perform the experiment. The RapidMiner as an Open Source Software was adopted for use to implement the SVM algorithm whereas the Clementine was used for the C5.0 decision tree algorithm. Chi-Squared statistics was used for the selection of top 30 terms, each class in the training dataset. Application was made on document frequency using Chi-Square in place of term frequency. Matrices for training and testing were formatted through the use of a Boolean representation which set the value 1 for term found in the document. And the value 0 for term who do not 70% and 30% sizes of the dataset were divided for training and testing respectively, in each corpus. The accuracy measure was used to determine the performances of the classifiers. The result indicated that the C5.0 algorithm performed more than the SVM. While the C5.0 had average accuracy of 78.42%, SVM had 68.65%. The disadvantage of C5.0 algorithm was that it was a block box algorithm that is only available in a commercial form.

Raho *et al,* (2016), investigated the effectiveness of using feature selection. The main objective of that study was to compare the performances between different classifiers in different situation through the use of feature selection combining stemming or not. The classifiers were: Decision tree, K-nearest neighbors (KNN), Naive Bayes method (NB), and Naive Bayes Multinomial Classifier (NBM). In order to embark on this study, BBC Arabic dataset carrying 4763 documents which has seven categories namely: Arts and culture – 122, Science and Technology – 232, Press World – 49, Sport – 219, Economy and business – 296, News of the world – 1489, News Middle East – 2356. There were 1,860,786 words and 106, 733 key word respectively in the dataset. There was also a pre-processing of the dataset. The classifiers were divided into two parts, for training and testing purposes. The performance of the classifiers aforementioned above were evaluated in terms of precision, accuracy, F-measures, recall, and time it took to build the model. Two feature selection methods namely: CHI and IG; and four classifiers as mentioned earlier were used , including the Waikato Environmental Knowledge Analysis (WEKA) tool, version 3.7 were used to determine the result of the experiment. From the result, it was indicated that Decision tree, Naive Bayesian method and Naive Bayes Multinomial performed better accuracy than K-nearest neighbor.

Othman & Al-Hamadi (2016), proposed a new algorithm known as the highest repetition of words in a text document (HRWiTD) for the classification of Automatic Arabic text. In order to carry out this experiment, data from Arab Source (Newswire) was obtained from the (SPA) Saudi Press Agency. It covered a sample of six categories. Availability of real classification for each text in corpus and availability of SPA texts on the web were the reasons for the choice of SPA as data source. The six categories were Culture, Economic, General, Political, Social, and Sports. The determination of the classification accuracy of the HRWiTD algorithm was done using confusion matrix method. The same dataset was applied in different classifier techniques that are famous. Models were developed depending on the use of C5.0, Naive Bayes, Support Vector Machine, Decision tree, and K-nearest neighbor classifiers. These models were created through the use of RapidMiner Software. The performances of the models were determined on the test set and the accuracy was evaluated using the cross-validation technique. The number of validation was set to X-Validation operators. The accuracy of the proposed algorithm (HRWiTD) and that of the techniques that were popular were compared. The popular classifiers had accuracy thus: C5.0 – 52.86%, KNN – 52.38%, SVM – 51.90%, NB – 51.90% and C4.5 – 30%. The best performance in classification process was achieved when the advanced methods used in term selection were: CHI, IG, None; For weight methods, we have Boolean, Entropy, Frequency, Relative Frequency, LTC, TFC, TFiDF. Two methods of sampling for selection of term namely: TF and DF were also used. It was found out from the study the best technique for use in the classification of Arabic texts in the domain selected was derived from HRWiTD algorithm. This was as a result of the fact that it had the highest accuracy compared to the popular classifiers.

Elhassan & Ahmed (2016), proposed Arabic text classification on full word. The main objective of the experiment was to explain and determine how effective, data preprocessing activities on full word in the accuracy of training model and classifier. The classifiers used in that experiment were the SMO (Sequential Minimal Optimization), NB (Naive Bayesian), J48, and KNN (K-nearest neighbors). These ones were used in building of the models for training. The proposed model was developed through the use of two approaches namely: the observation of data set involved, and the next being the elimination of stop words technique. To conduct the experiment, 750 documents were obtained from local newspaper called Ahir Lahza and Alyoum Altali and an international newspaper, Al-Raya, Asharq Al-Awsat, and Al-Hayat web sites between a periods of January 2001 to January 2015. There was a division of these documents into 5 categories namely: technology, sport, religion, political, and economy. Each of the categories has 150 documents, and the document being assigned to a category only. 105 are for every category used for the training of classifier while the rest were used for the testing. The text classifiers’ accuracy were compared in terms precision, recall, F-Measure. The method of evaluation was a cross validation model. The idea behind it has to do with the removal of some data before the commencement of the training set. This was used in the testing of the model’s performance. One of the cross validation was k-fold. It divided the data to k subset and a part used for the running test. In that experiment, 10-fold cross-validation was used in the evaluation of the model for training and data testing as well.

Duwairi (2016), proposed a distance-based classifier for categorizing Arabic text. The main objective of that study was to come up with a distance-based classification technique where categories are represented as feature vectors in an m-dimensional category space. In order to carry out the performance test on the classifier that was proposed, the Arabic text corpus were obtained from magazines and newspapers online. 1000 documents of different length and style of writing as well were taken. There were 10 predefined categories from these documents. There were 100 documents in every category. The predefined categories were: plants, internet, animals, religion, medicine, economic, art, technology, and politics. Two individuals carried out manually categorization of the documents that were collected. Only one category each was assigned to every document. As soon as a document was identified to fall under the ones with more categories, it was then pushed to the category that has maximal likelihood according to the judgment of human categorizer. 50 documents were randomized specifically for every category and applied in the training while the 50 that were remaining were to the testing operations. The accuracy of the classifier was determined in terms of precision, recall, error rate, and fallout. The results showed that the classifier that was proposed was very robust and accurate.

Al-Kabi *et al,* (2016), carried out an experiment on a topical classification of Quranic Arabic text. The main aim of the study was to evaluate the effectiveness of classification algorithms that were well known. These classifiers were: Decision Tree, (DT), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine (SVM). They were in the classification of different Quranic Ayats. That study had the following procedures as its framework: Arabic diacritics removal, Quranic symbols removal, the use of manual human topical in classifying Quranic āyāt (verses) for training and evaluation of the aforementioned classifiers. The study was dependent on the lexicon of Quran topics a authored by (Al-Khair and Kabbani, (2003), In that lexicon, the Quranic āyāt (verses) were classified into two 14 main categories and used in that study. They were in Arabic language and translated into English, and made up of the following: Part I: Pillars of Islam, Part II: faith, Part III: Science, Part IV: Working (MSI life), Part V: Call to God, Part VI: Jihad, Part VII: Human and Social relations, Part VIII: Ethical relations, Part IX: Regulation of financial regulations, Part X: Judicial relations, Part XI: Political relations, Chapter XII: Quranic stories, Chapter XIII: Previous religion, Chapter XIV: The diversity of divine discourse. There was the classification of each of these categories into sub categories. The first main topic which was “Pillars of Islam” was sub categorized into seven. They were in Arabic language sub categories but translated into English and namely: Chapter I: Unification, Chapter II: Religion, Chapter III: The Prophet (Muhammed, Peace be upon him), Chapter IV: Prayer, Chapter V: Fasting, Chapter VI: Zakat and alms, Chapter VI: Hajj and Umrah.

A total number of 6236 āyāt whole Quran, only 1227 āyāts (verses) which belongs to the selected classes were used in that study. Six measurements used in the evaluation of the performance namely: accuracy, recall, precision, F-measure, True positive (TP) rate, False positive (FP) rate, Receiver Operating characteristics (ROC). The selected three topical classes that were used include: Class I: all Quranic āyāt (verses) classified as ignorant of religion, Class II: has to do with the whole Quranic āyāt (verses) that were classified as oneness of God, and finally, Class III: concerns the whole Quranic āyāt (verses) that were classified as penalty of Apostates. The results from the experiment indicated that from the dataset used, Naive Bayes (NB) yielded more results that were accurate than the other classifiers, three in number.

Banu & Ganesh (2016), proposed a hybrid approach for an efficient classification using decision tree and support vector machine. That proposed model combined (SVM) Support Vector Machine and C4.5 together so that it will be possible for the achievement of a procedure that is integrated and efficient for use in classification. In that study, the method covered the division of all the data that were involved into two groups for the experiment as well as test data using a random manner in a 70 to 30 proportion. The data for the experiment were input into the standard Support Vector Machine (SVM), which lead to the estimation of the output. Using the SVM, classification of data was carried out which resulted to the obtaining of coefficients. New target was the name given to the class that was estimated. A calculation was made on the distance which arose between data that was derived at using the assistance given by support vectors, and also tallying with the class that was estimated thereby leading to the calculation of their value in average. The class that was estimated together with the distance value calculated in each of the data for the experiment serves as feature vector with the combination of the actual class of data served as the decision tree classifier input so that recalculation of the results will be made possible. In order to carry out an evaluation of that model (the SVM-DT), UCI repository was the source of 5 datasets whereas 1 was sourced from a Software termed StatSoft. The datasets for that experiment were: Leukemia, Ionosphere, Iris, Prima Indian, Wine, and Zoo. Each of them that have the following number of attributes namely: 4, 34, 4, 15, 14, and 18 respectively. The datasets were classified through the use of three classifiers such as SVM, C4.5 as well as the model that was proposed. The reason behind the use of these classifiers was for them to determine how the new model will behave in terms of performance. Cross-validations whose values were 10x10 was used in the estimation of the accuracies of the classification processes on the tree models. Waikato Environment Knowledge Analysis (WEKA) was the toolkit that was used in the experiment. In the WEKA was inbuilt C4.5 while STATISTICA was used for the classification of SVM. The result indicated that the proposed model performed more than other classifiers that were in terms of accuracy and time analysis.

Nidhi (2016), carried out an experiment on domain-based classification of Punjabi text documents using Ontology and hybrid based approach. The main objective of that study was to identify the best Punjabi text classifier suitable for Punjabi language. There were two algorithms new namely: the Ontology Based Classification, as well as the hybrid approach (these concern the combining of Ontology Based Classification and that of Naive Bayes. They were proposed for use in the Punjabi text classification. In order to classify Punjabi text classification, corpus used for that experiment had 180 Punjabi text documents, for training data, 45 files were used for testing. There were 3313 words in the training sets. These words were used for Punjabi text classifier training based on Centroid Based and Naive Bayes Based Techniques for classification. Sports related corpus made up the documents and was sourced from the new web sources of Punjabi such as ajitweekly.com, jagbani.com, and likhari.org. There were seven classes of unlabeled classified documents, namely: tennis, Olympic, kabbaddi. Others were: Hockey, football, cricket, and badminton. Implementation of the system was carried out through the use of C#.net platform. There were 2319 words in the stopword list that was manually prepared for the purposes of classification.

Arrays and files were the data structured for the classification of Punjabi text. Three experiments were conducted. Each class was in experiment 1 calculated. This calculation was for each classifier so as to determine the F-score, precision, and recall. Comparison was made between the classifiers in experiment 2. That experiment was dependent on documents that were not relevant and were obtained by each of the classifier from the total documents that were not relevant in the collection. The results in experiment 2 indicated that 2% only in the documents that were retrieved, were documents that were not relevant as hybrid approach and Ontology based were used in the classification of Punjabi text document. Whereas 5% and 6% for centroid based classifier as well as Naive Bayes respectively were non relevant documents that were retrieved. Average value for F1, Recall, and Precision in experiment 3 was carried out for each classifier. By comparing with other classifiers, Ontology Based classification performed better than Precision and Recall that recorded 89% and 85% respectively. From the results obtained from the experiments, it was concluded that Hybrid classification and Ontology Based Classification gave results that were better than that of standard classification algorithms like Naive Bayes and Centroid Based in Punjabi text classification.

Abutiheen *et al,* (2016), proposed an approach titled “Master-Slaves technique for the improvement of Arabic text classification”. That study involved two phases. The first phase involved the collection of Arabic corpus made up of 16757 text files. These collected text files were categorized into five classes manually. The second phase had four different classifiers which were implemented on the corpus that were collected. Classifiers used in that experiment were: Maximum weight (MW), Multinomial Logistic Regression (MLR), K-Nearest Neighbor (KNN), and Naive Bayes. Data for that experiment was the corpus that was built from a newspaper called “Al-Sabah”. There were 16757 Arabic documents that for the first time used in carrying out supervised machine learning. This corpus was manually segmented into five categories. There were two divisions of the dataset namely: training data with 90% allotment while the remaining 10% were for testing data. Implementation was carried out using naive Bayes classifier as Master while others served as slaves. Changing of the Naive Bayes classifiers (Master) probability was realized through the use of the results of these slave classifiers. Implementations of the four classifiers were done individually. Voting technique was also implemented as well as on the corpus collected so that the efficiency and effectiveness of the technique that was proposed will be checked. After the pre-processing of the Arabic text documents, all the tests were carried out. The pre-processing were tokenization, stemming, and stop-word removal. Vector weights were used to represent each document. 10-fold cross-validation was adopted for use to test results’ reliability. From the results, it was shown that the technique which was the Master-Slaves gave good result in terms of accuracy of the others.

Rasjid & Setiawan (2016), carried out a performance comparison and optimization of text document classification using K-NN and Naive Bayes classification techniques. That study was focused on classification of data through the use of two from the six approaches used for data classification. The classifiers used were K-Nearest Neighbors (KNN) and (NB) Naive Bayes. The objective of that study was to make a comparison on two documents for text classification technique so that an optimal value for K in K-nearest neighbor will be found or obtained. Data used to perform that experiment was obtained from TREC legal track carrying more than 3000 text documents, as well as more than 20 types’ classifications. Six (6) were chosen from these 20 types classification namely: Bankruptcy, Administrative Law, Trade Practice, Practice and Procedure, Migration, and Corporation. These documents were tagged by topics. K-nearest neighbor (K-NN) and NB (Naive Bayes) were the classification methods were used. 30% of the total data due for that study were used for training purposes. The classification of text was carried out through the use of K-nearest neighbor and Naive Bayes techniques. That was made possible as RapidMiner Software was used, and 1-25 with odd integers only, were chosen as the value of K. The reason for the choice of odd number was to avoid a case in which positive and negative result will be arriving at with equal number. The performance of that experiment was determined using Accuracy, F-measure, Precision and Recall. From the result which was arrived at using F-measure, it indicated that K-nearest neighbor performed more than Naive Bayes classifier.

Faidi *et al,* (2016), compared Arabic natural language processing tools for Hadith classification. The main objectives of that study were to make comparisons on how different techniques to classify Al-Hadith Al-Shareef will perform. Six (6) Arabic tools were used to carry out the analysis. They were: Al Stem Darwish, Al-Stem Alex Quadrigrams, Khoja’s Stemmer, Trigrams and AraMorph. DT (decision trees), NB (Naive Bayes), and (SVM) Support Vector Machines algorithm. TF-IDF was used in the computation of each word’s relative frequency in a particular document. While cross validation was used in the evaluation of the classifiers’ result. The dataset for that experiment was made up of hadiths obtained from Sahih Al-Bukhari which contains the traditions of Prophet of Islam Muhammad’s (PBUH). A hadith shows all that the prophet said and did. There were two branches of it namely: the Sanad meaning chain narrators and the Metn meaning what the hadiths contains. The term “book” was used by Al-Bukari to classify Al-Hadith’s subject. There was a class, a category and a chapter, Sahih was splitted into 7031 hadiths according to their subjects. 795 Hadiths were selected. While 23 categories wre divided from them. They were: The Book of Prayer Hall, The Book of the Eclipse Prayer, The Book of Oppressions, The Book of Bathing, The Book of Menstrual Periods, The Book of the Two Festivals, The Book of Manumission of Slaves, The Book of Distribution of Water, The Book of Agriculture, The Book of Wills and Testaments, The Book of Patients, The Book of Al-Adha Festival Sacrifice, The Book of Virtues of Madinah, The Book of Penalty of Hunting while on Pilgrimage, The Book of Minor Pilgrimage, The Book of Actions while Praying, The Book of Invoking Allah for Rain, The Book of Shortening the Prayers, The Book of Hiring, The Book of Loans, Payment of Loans, Freezing of Property, Bankruptcy, The Book of Divine Will, The Book of Tricks, and the Book of Supporting the Family. The data was classified and results obtained from each of the algorithm. This was done so as to determine the accuracy shown by each of the classifier. Comparisons were carried out between the classifiers in the performance of the accuracy so that the best among them will be identified. The accuracy of these classifiers was shown through recall, F-measure and precision averages. From the result, it was shown that Sequential Minimal Optimization (SMO) had the highest accuracy followed by NB (Naive Bayes) and next was J48 (C4.5) classifier. Conversely, the result showed that Khoja’s Stemmer performed more than other tools. Meanwhile that experiment was performed through the use of Waikato Environment Knowledge Analysis (WEKA) tool kit. The performance and efficiency of the following were determined using that tool Kit. They were: Sequential Minimal Optimization (SMO), C4.5 algorithm implemented using the name J48 Algorithm and Nave Bayes.

Ahmed (2016), carried out a study to determine impact of text pre-processing and totally different term weighting schemes on Al-Hadith Al-Shareef Classification. In addition to that, presentation and comparison were made in a bid to determine how effective three distinct automatic learning algorithms were in classification of Al- hadith Al-Shareef in eight selected books which depends on Sahih Muslim. The eight books were: the book of Fasting, the book of Faith, the book of Knowledge, the book of Hajj, the book of Eclipse Prayer, the book of Obligatory Charity the book of Good manners and the book of Praying. The classifiers that were used during that process of classification were: Naive Bayes (NB), Support Vector Machine (SVM), and Complement Naive Bayes (CNB). 10-fold cross-validation was carried. Technique such as Term Frequency-Inverse Document Frequency (TF-IDF), (TF) Term Frequency, (TO) Term Occurrences, and Binary Term Occurrences (BTO) were used in the computation of relative frequency for every word found in a particular document. The corpus was shared into two (90% for training while the rest 10% was used for testing) through the use of stratified sampling. The results showed that stemming and pruning, normalization of document, and term weighting reduced dimensionality in a dramatic way, giving room for text representation and having text mining performance impacted directly. The results further indicated that Complement Naive Bayes (CNB) performed more than others, such as Naive Bayes (NB), and Support Vector Machines (SVMs).

Olayah & Alromima (2016), proposed an Automatic Machine Learning Techniques (AMLT) for classification of Arabic text documents through the use of term collocations. Arabic documents were the source collocations. The collocations that were obtained from the Arabic documents were divided into categories of four (Science, Religion, Politics, and Economy/Business). These were divided further into 200 documents each totaling 800 datasets. The experiment was conducted through the use of Waikato Environment for Knowledge Acquisition (WEKA) tool kit. The results of the proposed approach were compared with the use of the following classifiers: Support Vector Machines (SVM), Naive Bayes (NB), J48, and K-Nearest Neighbor (KNN). This was done to determine if the classifier has more accuracy for Arabic text which was dependent on term collocation. It was indicated that from the results, the proposed technique performed more than others.

Arjaria *et al,* (2016), carried out a study on Two Phase K-Nearest Neighbors Approach. The aim of that study was to present a new idea which will be used in dealing with the classification in two phases. While the first phase dealt with the bringing out of information that was useful from the training space which concerns how the each training sample behaved in the neighborhood list of other samples for training. This behavior that occurred brought about the decision concerning each training sample to be among the three classes which include neutral, unimportant, and important. The second phase had a rearrangement of the information that was obtained from the training samples which is nearer to the testing sample. This arrangement was carried out through the removal of the samples that are unimportant. The classification decision removed completely the training samples that are unimportant and take the important as well as the neutral class that form the samples for training. An algorithm was designed so that provision will be made of extra weights which are important samples considering their position in the list of neighbors, its frequency occurred as a neighbor of other samples for training and the number of those samples for training within that class that was used while undergoing the training. Testing of the performances was carried out on three databases, seven categories of Reuters-21578 that are most in terms of frequency, seven categories of TDT2 most frequently occurred corpus. From the results of the experiment, it was found out that the proposed Two Phase K-Nearest Neighbors Approach performed more than the traditional K-Nearest Neighbors approach.

Khreisat (2016), carried out an experiment to classify Arabic text documents through the use of N-gram frequency Statistics technique which employed a dissimilarity measure known as “Manhattan distance”, and Dice’s measure of similarity. These were carried out for Classification purposes. Corpus that was made up of Arabic documents was obtained through the use of Arabic news articles that were obtained online websites of so many Arabic newspapers. The corpus was in four categories of text documents namely: weather, technology, economy, and sports. Documents from sports and economy were larger and ranges from between 2KB and 15KB, and between 2KB and 18KB for sports and economy respectively. The smaller document has 2% of the entire number of documents found in the category of economy and sports. These documents were subjected to text preprocessing. For training classes, 40% were chosen from the corpus while 60% were chosen from the corpus for testing of the procedure for classification. Both the training documents and the ones for classification passed through the same process. Each of the document for classification were subjected to text preprocessing phase, after that, was the generation of N-gram profile. There was a comparison of the N-gram profile of each text document against the profile of the entire documents in the classes for training considering their similarity. Two measures were adopted for use. The first one was the dissimilarity or distance, known as “Manhattan distance,” while the second one was the Dice measure of similarity. For the purposes of comparison, the Dice measure was used. Results indicated that N-gram text classification through the use of Dice measure performed more than Manhattan measure.

Kanaan *et al,* (2017), carried out a study by comparing text classification techniques applied to Arabic text. The main of that study was the implementation of three automatic text classification techniques for Arabic text. The classifiers that were used in that experiment were naive Bayes, Rocchio, and K-nearest neighbor algorithms. Because Arabic corpus was not available publicly, in-house corpus obtained from Arabic newspaper archives online was used for testing. These newspapers were Al-Jazeera, Al-Hayat, Al-Dostor, Al-Nahar, and Al-Ahram, and other few specialized web sites. 1445 documents served as the data set. They were of nine categories and different lengths. Each document was kept in a different file within the directory of the category. Preprocessing activities were carried out. From the results, it was indicated that Naive Bayes performed more than K-nearest neighbor and Rocchio.

Saleh (2017), carried out a study on automated Arabic text categorization using SVM (Support Vector Machines) and NB (Naive Bayes). The aim of that study was to investigate SVM and NB approaches or techniques on various Arabic data set. The reason for that comparison was to determine the evaluation measures that were popular than others. The experiments were implemented and it was discovered from the result that SVM performed better than NB considering all the measures that were used.

Azam *et al,* (2017), carried out a study on feature extraction-based text classification using K-nearest neighbor algorithm. The main objective of that research was to make an analysis on the performance of classification algorithms through the use of Scopus dataset. The dataset belongs to Elsevier, they also maintain them. The data set were obtained online by subscribing. There were 22k titles obtained from more than 5k publishers, while 20k was from written journals and reviewed by experts from different scientific, social sciences, medical and technical fields. For the purposes of analysis, extracts were made from abstracts having 10k documents got from Scopus with their categories of journals. The categories were five in number namely: Medicine, Agricultural and Biological Sciences, Finance, Engineering, and Mathematics. Preprocessing activities were data preprocessing, data training, and testing of different algorithms for classification through the use of rapid miner. In that study, the classifiers were K-nearest neighbor (K-NN) and Naive Bayes (NB). The results of the experiments indicated that K-Nearest Neighbor outperformed more than Naive Bayes.

Patel & Soni (2017), carried out an experiment by increasing accuracy of K-Nearest Neighbor classifier for text classification. In order to carry out that experiment, inverse cosine distance weighted voting function was proposed for classification of text. The data set for that study were the Reuter-21578 and 20 newspaper group text data set which had 21578 documents and 65 categories. 5 top categories such as: Earn, Crude, Money, Acq, and Trade were selected for use in the elevation of the algorithm. Among 61188 attributes, 1400 were selected after using dimension reduction method. For testing process, some data set were randomly chosen. They were: 1413, 1250, 1000, 700, and 500 documents for training. While the 20 news groups’ data set had 20 categories, documents of 18774 and attributes of 61188. 5 top categories such as: Atheism, Motorcycles, Space, For sale, and Electronics etc. were selected. The data set was changed to tf-idf format for use in further processing. 3000 top attributes were used with 10 documents for the purpose of testing while for training, 1055 were used in the carrying of the experiments. The proposed classifier was compared to simple K- Nearest Neighbor. The results of the experiment indicated that the proposed weight function with K-NN classifier outperformed more than simple K-NN.

Taeho (2017), proposed the modified K-nearest neighbor (K-NN) algorithm that will consider the feature similarity and also applied to text classification (categorization), in a study titled, “Text classification using features similarity-based K-Nearest Neighbor. Data set were collected from NewsPage.com. Four categories were predefined in that collection, texts were obtained from the collection depending on the category. There was a classification of each text exclusively in one among four categories. The four categories that were predefined were: Sports, Internet, Health, and Business. There was a random selection of 375 texts in each category. The 375 set of data were partitioned into 300 texts for training and 75 texts for testing. Opinosis was also used in the experiment. Three categories of Hotel, Electronics, and Car were predefined. And texts were used to evaluate the text classification approaches. 51 text labeled with one of these categories of three (Hotel, Electronics, and Car) were encoded into vectors that are numerical in nature having input sizes thus: 200, 100, 50, and 10. 33 training examples were selected for training through the use of KNN version of computation so as to arrive at the similarities. From the results it was found that the proposed version performed more than the traditional versions.

Bajeh *et al,* (2017), carried out an experiment for performance evaluation of some selected machine learning algorithm for sentiment analysis. The aim of that study was to present a comparative analysis of the performance of naive Bayes, k-nearest neighbor, and support vector machine in sentiment analysis. Tweets serving as text data were obtained from Twitter Platform through the use of its tweety API. There were the airline, and Deflate-gate datasets. They were tweets arising from airline services and that of 2015 super bowl show respectively. Data set from the airline were made up of 14,640 tweets while the deflate-gate was 11,814 used for the purpose of analysis. Data pre-processing took place through the use of methods such as tokenization, stop word removal, and stemming. Transformation of data from unstructured to structured format took place through the use of these methods. This made it easier for the classification algorithm to learn. Success was attained in the preprocessing as Sckit-Learn’s Hashing Vectorizer was used, a python API that changes a text collection documents to a token occurrences in a matrix form. From the result it was indicated that KNN performed more than NB and SVM.

Shilpa *et al,* (2017), carried out a comparative study on data classification algorithms KNN (K-nearest neighbor) and SVM (Support Vector Machine) in diagnosing heart disease. The study investigated the application of KNN and SVM in the helping of healthcare professionals to diagnose patients that have heart disease. 3 different secondary datasets were obtained for machine learning data repository of UCI. They were implemented on the algorithms so that the accuracy of the data that were obtained will be identified. From the results of the experiment, it was indicated that SVM achieved accuracy that was higher and more efficient than KNN.

Khamis *et al,* (2017), studied on the application of K-nearest neighbor classification in the medical data mining. The objectives of that study were to: evaluate to what extent K-nearest neighbor classifier enhance efficiency and accuracy amongst patients seeking emergency treatment in Kenya, evaluate the factors affecting the implementation of K-nearest neighbor mining technique in Kenyan hospitals, design a repository with efficiently classified data for easy data mining. That study was carried out in Coast Province with one referral hospital (Coast general hospital) and other hospitals within the seven districts namely: Tana River district hospital, Taita Taveta district hospital, Mombasa district hospital, Malindi district hospital, Lamu district hospital, Kwale district hospital, and Kilifi district hospital. Questionnaire was used to elicit responses from the respondents. It was made up of two parts. While Part A was concerned with the collection of demographic data from the respondents, Part B was concerned about the evaluation of the extent K-nearest neighbor classifier could enhance the delivery of service to patents who are looking for emergency treatment in the country of Kenyan. And Part C was for the evaluation of the factors that affect the carrying out of K-nearest neighbor mining method in the hospitals in Kenya and other services that may be readily available. Descriptive statistics of mean, standard deviation, and variance were used to analyze the quantitative data like the demographic characteristics about the respondents obtained through the use of Part A questionnaire. Those data were then shown through the use tables for frequency distribution. It was found from the results that considering objective 1, to: evaluate to what extent K-nearest neighbor classifier enhanced efficiency and accuracy amongst patients seeking emergency treatment in Kenya, using KNN algorithm can help greatly in the reduction of errors in diagnosis, reduction in time spent on diagnosis whilst making improvement in the effectiveness and efficiency in terms of treatment. For objective 2 which was to evaluate the factors affecting the implementation of K-Nearest Neighbour mining approach in Kenyan hospitals, it was indicated from the results that KNN mining approach was affected mostly by costs in administration and efficiency of classification.

Hassanat *et al,* (2017), researched on solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. That study proposed a new solution for choosing the K parameter in the KNN (K-nearest neighbor) algorithm. The solution relied on the information that was obtained from ensemble learning whereby a K-nearest neighbor classifier that was weak, was used every time with a different K. The application and comparison of other methods were made using the proposed classifier. As the experiments were on, 28 different data sets were chosen from the machine learning repository of UCI so as to make a representation of real life problem of classification. The data sets used were: heart, balance, caner, German, liver, vehicle, vote, BCW, Haberman, letter, recognition, wholesale, Australian, Glass, Sonar, Wine, EEG, Parkinson, Iris, Diabetes, Monkey Ionosphere, Phoneme, Segmen, Vowel, Wave 21, Wave 40, Banknote, and QSAR. Each of the dataset was divided into two sub data sets, one was for training while the other was used in the testing. While 30% were for testing, the rest were for training. Design was made often types of classifiers so that their performances will be compared with the classifier that was proposed. These classifiers that were designed were: IINC, 60 – NN, 45-NN, 30 – NN, Vn – NN, 9- NN, 7 – NN, 5 – NN, 3 – NN, and 1 – NN. The traditional K-nearest neighbor (KNN) classifier that use small, medium and large number of neighbors were included, IINC classifier in addition, which serves as the best state-of-the-art classifiers. Comparisons were made between the proposed classifier and others through the use of a group of experiments found in life problems that are real. The results of that experiment indicated that the classifier that was proposed outperformed the traditional KNN classifier which adopted the use of different number of neighbors.

Zhang *et al,* (2017), proposed a K Tree method for learning different optimal K values for test/new samples that are different through the involvement of a training stage in the KNN classification, while carrying out a research on efficient KNN classification with different members of nearest neighbors. Data sets obtained from UCI Repository of Machine learning were used for the experiment. These data sets were 20 in number, for the evaluation of the methods that were proposed and that of the competing ones on the task of classification, considering the accuracy of the classification and the running cost as well. These data sets of different types are: low-dimensional, high-dimensional, binary, multiclass, and unbalance data sets. They were used for the evaluation of the robustness of the methods that were proposed. Some of these data sets were used, namely: Abalone, Blood, Balance, Climate, Australia, Car, and German. These ones were used to carry out the experiments having different size in terms of sample. The other data sets were used in the experimentation of feature of different numbers. They include: Madelon, CNAE, Hill, Db-world, LSVT, Arcene, Gisette, Libras. From both group of data sets, climate had 46 samples in the positive form, while 494 samples belong to the negative. German data set had 700 samples from the positive parlance while 300 samples were from the negative. These can be taken as data sets that are in an imbalance form. Tenfold methods of cross validation were employed on the methods. The entire data sets were partitioned into ten subsets randomly. One subset was then selected for use in testing while the nine remaining subsets were used for training. The whole processes were repeated ten times so that there will no bias when the process of partitioning data set for cross-validation is going on.

In that study, the following state-of-the-art methods were selected as the methods for computing. They were: KNN, KNN-based applicability domain approaches (AD-KNN), KNN method that was based on sparse learning (S-KNN), (GS-KNN), filter attribute. From the results of the experiments, it was found that the two new KNN classification algorithms that were proposed, i.e., the K Tree and the K\* Tree methods, outperformed the methods for competing considering the accuracy of classification and the cost of running.

Nikhath *et al,* (2017), researched on Building a K-nearest neighbor classifier for text categorization. The aim of that study was the introduction of an e-mail classification application text categorization through the use of K-Nearest Neighbor (KNN) classification. There was the preprocessing of the documents that were found in the training set. They were represented through the use of model based vector space made up of features that were taken from each of them. The Euclidean distance measured was adopted for use in the measurement similarity that exist documents that are two in number. The traditional precision and recall were used to determine how effective the classification is. It was found out from the results that K-nearest neighbor classifier was reliable for text classification.

Yingbo *et al,* (2017), proposed a teaching resource classification method based on the improved K-Nearest Neighbor (K-NN) algorithm in their study titled “Classification Method of Teaching Resources based on improved KNN Algorithm. Text class for primary and secondary school resources for teaching served as the object for the research. It was combined with the characteristics of the domain, and there was an improvement in K-nearest neighbor algorithm. A self – built corpus was a data set for the experiment. There were 3,000 texts in the corpus. These texts were of different types and not distributed evenly. Categories that involved totaled 10. They were: Mathematics, English, Biology, Chemistry, Physics, History, Geography, Politics, Chinese, and others. There were 2,100 texts and 900 for training and test, respectively, among these categories. Due to the open source code that Weka Platform made available, implementation of the second development was made, the improved algorithm was carried out. Testing of its performance was made improvement in the algorithm was compared and analyzed after the effect of the classification was observed. The results from the experiment indicated that method of classification of primary and secondary school resources for teaching based on K-nearest neighbor algorithm was effective and feasible.

Salvador-Meneses *et al* (2017), proposed a variation of the K-nearest neighbor (KNN) algorithm that will work with categorical data that was compressed, in a study titled compressed KNN: K-Nearest Neighbors with data compression. In that study, two datasets were used for the test. They were: the Census Income Data Set (CIDS) that gives a representation of data set that are mixed and has both numerical and categorical variables, and the other was Winconsin breast cancer (original) (WBC – original) that gives a representation of data set that have categorical variables only. The census income data set has 32,561 observations while Wisconsin Breast Cancer (original) had 699 observations. They have 15 and 11 attributes, respectively. There were experiments and tests using the WBC and CID data sets. The results indicated that the proposed methodology carried out a reduction in the amount of memory that the K-nearest neighbor algorithm used thereby keeping the same error in the percentage level of the classification that was made using the original method.

Fanny *et al,* (2017), compared a news categorization using K-nearest neighbor, Naïve Bayes and support vector machine in a study titled a comparison of text classification methods K-NN, Naïve, Bayes, and support vector machine for News classification. The methodology used in classifying the documents (news) was divided into two processes, the document for processing and the other for validation of the classification. Prior to the process of validation, processing of the document has already been done so that some words or characters that are not important will be removed, doing this will make the classification process to be ease. Four steps were involved in the documents processing, namely: tokenization, removal of stop-words, removal of tokens (through length and stemming).

That experiment used documents which were obtained from the following news areas such as BBC news, New York Times, Fox News, and ABC news. The documents were shared into categories of 6 namely: politic, technology, health, sports, entertainment, and economy/business. Total data for each category was 30. Processing of the document and the process of validation took place. After these was the splitting of data for training and testing purposes. Three experiments were carried out in that student. From the results of the experiments, it was found that K-NN performed more that Naïve Bayes and SVM classifiers.

Nur *et al,* (2017), carried out a study on the implementation of text running classification as model in the conclusion of Tafsir Bil Ma’tsur and Bil RA’YI contents. The purpose of that study was for the building of an application that carries out the responsibility to classify type of interpretation in an automatic manner into two classes for users’ convenience in the Al-Qur-NN. These classes are Tafsir Bil Ma’tsur and Bil RA’YI. K-nearest neighbor (KNN) was used as the classifier. The methodology that was adopted in that study was divided into several stages. They were: data preprocessing, data processing, and comparison of the algorithms that appeared best and carrying out analyzes of the results that appeared best as well. Swrah Al-Baqoroh’s found in Al-Misbah M. Quraish Shihab’s Tafsir of the Quran served as the textual data which was processed to bring about information was used as the material for that study. In the same manner, Tafsir Bil Ma’tsur and Tafsir Bil RA’ YI were the main class that was interpreted in the process of classification with the use of K-NN, MK-NN, and FK-NN (Classifiers) algorithm. That study commenced from the collection of data, namely changing of the letter Al-Baqoroh, text simplification, preprocessing and making comparisons of the algorithms that were used, which were three in number. Clearing process took place and the text was changed to lowercase (case fold), all the non-letter characters were removed, and punctuations deleted. Texts in documents1 – 286 which were the data, from that stage can be possibly used. At this juncture, 10 experiments were carried out and the classification of the three algorithms namely K-NN, MK-NN, and FK-NN on cross validation. From the results of the experiments, it was found that K-NN performed more than other with high accuracy.

Elnazee *et al,* (2017), carried out a study to evaluate the performance of (K-NN) K-nearest neighbor algorithms while determining the learning styles of students. K-nearest neighbor (K-NN) algorithm or classifier was used in the learning styles of students’ domain in the predicting of the outcomes that are possible through the use of different numbers of records obtained from the data set, the ratio that exist in data set for testing and that of the data set for training; and the K value as well. the reason for doing this was for the minimization of the classification that was missed in other machine learning algorithms. Felder Silverman Questionnaires was adopted for use in the study for collection of data. The questionnaires were given to 200 respondents for the purpose of eliciting responses. Selection of these respondents was made from the Faculty of Computer Science and Information Technology, Bayero University, Kano, Nigeria. Use case diagram was used as well. The system was made up of the administrator and distributor. While the distributor gives the questionnaires that were generated through the use of Felder Silverman System. They as well received the questionnaires, logged them into the system with the use of student ID as the questionnaires stated. They were taken from the questionnaire. In order ascertain the learning style of the student, the Felder Silverman algorithm was run. As that was concluded, the learning style of the student was then recorded and stored in the database. To predict the learning style of the student the administrator logs in and from the knowledge base gets the data and run the K-nearest neighbor (K-NN) algorithm or classifier. Evaluation was made on the learning style of the student which was predicted by the K-nearest neighbor classifier, and comparison as well with one that was they had through the use of Felder Silverman system of classification. Evaluation was made on how the K-nearest neighbor classifier performed as the predicting of the learning of student was made. The results from the study indicated that K-nearest neighbor classifier performed creditably and found to be good for data running.

Nedungadi *et al,* (2017), proposed a new hybrid algorithm which used Principal Component Analysis (PCA) and K-nearest neighbor for text classification, but carried out K-NN using a little set of neighbors rather than complete data vectors in the space that was projected, thereby making the computational complexity to be reduced. Synthetic and real dataset taken from Reuters data set having 6532 documents, features ranging up to 5741 that belong to categories, up to 52 were used in that study. The experiments were performed. The results indicated that the hybrid model was able to classify data with a level of accuracy that was very high as a little number of principal components and a set of projected data were small for K-NN. This brought a reduction in time of computation.

Mustakin (2017), carried out a study on the effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbor classification. The experiments were carried out through the use of K-means clustering against (K-NN) K-nearest neighbor algorithm whose validation was made by confusion matrix. Collection of data was made through the use EMIS data obtained from the Ministry of Religious, of Pekanbaru city and number that was generated randomly. Transformation step has to do with the use of numerical scale that was based on each attribute. Data Distribution was carried out in 4 fold of 4900 data before clearing was carried out. A comparison was made of the data through random generation of 150,000 data. While 112,500 were for training, the testing data were 37,500. Distribution of data through K-fold cross validation will be compared to techniques for data distribution through the use of K-means clustering. These were the division of data into 7,6,5,4, and 3 group sections. 30% data were used. Each of the group of data for testing and training were merged. K-nearest neighbor in conjunction with confusion matrix were used for the implementation of the two data distribution models. 9 (nine) attributes for the process of classification were: Average Parents Earnings, Mothers Job, Mothers Formal Education, Father’s Job, Father’s Formal Education, Class level of Education, Category Students Study, and Gender, while the class targets of K-nearest neighbor were Level of Family Education Low, Medium and High and Economic Life. All of these make up the Educational Management Information System (EMIS). A random data was also implemented. It was found out from the results of the experiments that the use of K-means clustering against (K-NN) K-Nearest – Neighbour Validated High the use confusion matrix has the highest accuracy of 93.4% than the K-Fold Cross Validation data distribution technique.

Wang *et al,* (2017), proposed a novel multi-label classification algorithm that was based on the random walk graph and the K-nearest neighbor algorithm (captioned MLRWKNN), in a study titled “a novel multi-label classification algorithm based on K-nearest neighbor and random walk”. The researchers used six data sets in the conduct of that experiments and evaluation of the Multi-Label Classification (MLC) algorithms. These data sets were Medianmill, Yeast, Scene, Medical, Genbase, and Flags. Application domains like video data, biologic data, images and texts were covered while the number of labels was from 6 to 101 and also vary. Seven state-of-the-art Multi-Label Classification (MLC) algorithms were selected for use to compare the performance. They were: Binary Relevance KNN (BRKNN), Hierarchy of Multi-Label Classifiers (HOMER), RAndom k- LabElsets (RAKEL), Rank Support Vector Machine (Rank – SVM), Caliberated Label Ranking (CLR), Label Powerset (LP), and Multi-Label KNN (MLKNN). Among these, BRKNN, HOMER, RAKEL, CLR, and LP were of the problem transformation methods (PTMs) category while Rank-SVM and MLKNN belonged to Algorithm adaptation methods (AAMs). The study adopted 10-Fold cross validation while an average of 10 experimental values was considered for the purposes of comparisons.

The proposed multi-label classification algorithm based on the random walk graph for the K-nearest neighbor (MLRWKNN) was concerned about the problem of ranking; there were four ranking-based criteria to determine the validity of the method. These ranking-based criteria were: Ranking Loss (RL), Coverage (COVE), One Error (OE), and Average Precision (AP). The experiments were carried out, and the results indicated that the proposed MLRWKNN algorithm had result that was better than other seven MLC state-of the-art algorithms.

Adawuofor, (2018), presented an improved classification model for Igbo text using N-gram and K-Nearest Neighbor approaches. Word-based N-gram model was used for text representation in that study. The classification of the text was carried out on the three N-grams represented texts using the K-Nearest Neighbor model based on similarity measurement so as to test the effectiveness of the model. Object-Oriented design methodology was used in the research. The Python Programming Language was in conjunction with the Natural Language Toolkit (NLTK) used in the implementation. The performance of the Igbo text classification system was measured through the computation of the precision, recall, and F-1 measure of the result got on Unigram, Bigram, and Trigram represented text. The Igbo text classification on bigram represented text had highest degree of exactness (precision). The result obtained with three N-gram models indicated the same level of completeness (recall) while trigram had the lowest level of precision. This indicated that the classification on bigram Igbo represented text performed better than unigram and trigram represented texts. Therefore, considering the aforementioned, bigram text representation model was highly recommended for any intelligent text-based system in Igbo language.

Mowafy *et al,* (2018), presented a model for text classification that shows level of phases by building automatic text document classifier. The relationship existing them were presented. To evaluate the proposed model, 20 news group dataset were used to carry out the experiment. The results of that experiment indicated the superiority in the use of multinomial naive Bayes combining TFIDF than K-nearest neighbor as text document classification approach. The model that was proposed produced results that are better, improvement in the performance of the classification process, and confirmation of their compatibility concept existing between the various phase’s techniques that were selected. Finally, in spite of the fact that the performance of the model that was proposed was better, based on the embedded techniques that were selected, the performance of other techniques can be improved as can be seen in the results of the experiment. Since improvement in the KNN-TFIDF’s accuracy can be possible through the use of the model that was proposed to 76% from 71%. The proposed can therefore be extended to other approaches for use. A suggestion was made that the proposed model should be upgraded through the study of other classifications techniques that are compatible with techniques for feature selection and different term weighting schemas so that the performance will be better.

Mandral & Sen (2018), compared then efficiency of four methods that can be used for categorizing Bangla web documents. The four methods were Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Decision Tree (C4.5). Empirically, the classification performances were compared using Bangla text documents. The researchers came up with their own developed corpora with the name BD Corpora. A feature extraction tool was used and selection was made. The findings of the experiments showed that on training sets that are well – organized and small, NB and KNN have higher capability comparing SVM and DT (C4.5) while carrying out documents categorization. But then, in a situation where the document is larger, SVM is better than others. While comparing the time for training, all the classifiers do not pass the same time for learning. More than what is taken for training by DT (C4.5) than the other three algorithms except SVM whose learning rate is quick. Again, from the result, average – F-measures showed SVM as the classifier that has best result next was NB, DT (4.5) as well as KNN. Furthermore, it also shown that two classifiers are very sensitive to the parameters that was set. These classifiers were SVM and KNN. That is to say that parameter that is different will give a result that is different. Because of this, defining parameter settings that are best is very important.

Basma et al,(2014), proposed a study on an Independent Living for Persons with Disabilities and Elderly People using Smart Home Technology. The study took a snapshot on the state of the art in the smart home technology for elders and disabilities people. Also it proposed a new Elders/Disabilities Wireless Smart Home for assistive independent living (E/D-WSH) with its approximate cost compared with the Indian home automation market. Their proposed system was not backed up with real-time data hardware devices for the discussed smart home technology.

Ismail & Erkan (2017). implemented a smart house application using wireless sensor networks. In the study, a smart home via smart phones and pcs was explained. In addition, the control of electrical and electronic devices in houses and buildings was also scrutinized by the authors. The authors did a good job. However, performance evaluation of their developed model showed some routing deficiencies.

Toshifumi (2015), presented a study on In-home Health monitoring system for solitary elderly. In the study, the authors proposed a sensor-based monitoring system that evaluates the health status of solitary elderly based on daily living activities, and provides forecasts of emergency situations to a local nursing center without explicit user interaction. In addition, the authors also focused on three daily living activities: urination, kitchen work, and activities related to maintaining physical cleanliness because these activities are closely involved in maintaining a healthy lifestyle and are usually accompanied by the usage of tap water. The authors did a good job. However, the implementation carried out by them was only limited to few trials.

Heetae et al,(2018), proposed a study on IoT Smart Home Adoption. The study examined the smart home service features that current users require and empirically evaluates the relationship between the critical factors and the adoption behavior with 216 samples from Korea. The moderating effect of personal characteristics on behavior is also tested. The results of the analysis provide various theoretical and practical implications. However, the obtained result from the study was deficient in cost benefit analysis.

Salman et al, (2011), looked at Real Time Algorithm for the Smart Home Automation based on the Internet of Things. The authors proposed an algorithm for smart home automation system based on IoT using sensor nodes which are directly connected to Arduino Nano. The algorithm performs some basic local functions such as; Turning ON/OFF the lights based on the motion sensor and generating the alarm based on the gas sensor. The authors did a good job but could not further implement with IoT using sensor nodes which are directly connected to Arduino Nano

Dagmawi et al, (2019), presented a study on Reasoning in Multi-agent based smart phones. The study examined literature on home systems and provides a comprehensive overview of the essential requirements, assumptions, strengths, limitations, challenges and future research directions of their proposed reasoning systems. However, the discussed findings from the study were not implemented with a machine learning oriented model.

Akshaya et al, (2018), looked at Smart Pill Box. The study proposed a mobile application called smart pill box which addresses medical issues of the elderly. Smart pill box alerts the patient by giving an alarm signal to take the medicines in the correct time. The smart pill box also periodically checks the pill count and sends a SMS or notification to the caretaker or the pharmacy to refill the pillbox if the pills are about to get empty. The authors did a good job. However, comparative analysis of the Smart Pill Box shows that every user must possess a smart phone without which they cannot get pill notifications and communicate with pharmacy or their caretaker.

Sumit et al, (2017), proposed a study on Smart Homes for Elderly Healthcare. The study presented a comprehensive review on the state-of-the-art research and development in smart home based remote healthcare technologies. The reviewed study on smart homes for the elderly healthcare was not implemented to a model for more clarification and understanding.

Quynh et al, (2012), looked at Smart Homes for Older People. The study examined the concept of smart homes in a technologically driven society. Furthermore, the study discussion was focused on challenges in the use of smart homes among older people such as accessibility and ethical issues. The authors did a good job. However, their proposed system was not backed up with real-time data hardware devices for the discussed smart home technology.

Ieva et al, (2019), presented Deploying Smart Program Understanding on a Large Code Base. They argued that new developers face the task of classifying a huge number of files and functions without any help. Their proposed approach to this problem called FEAT automatically extracts topoi from source code by using hierarchical agglomerative clustering. Although the testing revealed that the approach was suitable to understand open-source software projects of size approaching 2,000 functions and 150 files, which opened the door for its deployment in the open-source community. But they could not use semantic classification of all projects source codes in the automatic detection of semantic clones at a very large scale as well as mining of test script to find hidden links between test cases.

Ieva et al, (2018), presented discovering program topoi through clustering. They stated that organizing the code in a more abstract way, closer to humans is an attempt that received interest from the software engineering community but unfortunately, there is no recognized recipe or tool that can concretely provide any help in dealing with large software repositories. So they propose an effective solution to this problem by automatically extracting program topoi, which are ordered lists of function names associated with an index of relevant words. Their research spans the domain of program understanding focusing mainly on feature extraction; automatically discover the main characteristics of a software system by analyzing source code or other artifacts. However, a major limitation of their study is that the developed model was deficient in benchmarking and cost benefits analysis.

Niazi & Hussan (2011), designed agent-based computing from multi-agent systems to agent-based models; a visual survey. The study used Scientometric analysis to analyze all sub-domains of agent-based computing. Their proposed system also employed a combination of two applications for analysis of Network workbench and Cite space where network workbench allowed the analysis of complex network aspects of domain, detailed visualization using CiteSpace. They did a good job. But could not include a detailed analysis of each of the sub-domains to further explain the Scientometric indicators inside each of multi-agent systems, agent-oriented software engineering and agent-based modeling.

Ananda & GU (2013), demonstrated multi-agent based protection on highly denominated distributed energy resources. The study analyzed the use of MAS approach on radial distribution system protection using dispersed adaptive rule-based protection supported by distributed database agent. The authors did a good job. But they failed to evaluate the study simulation results based on Matlab-Simulink with other algorithms for energy distributed resources.

Leitao et al, (2013), designed parallelizing multi-agent systems for High Performance Computing (HPC). The study discussed the parallelizing of MAS solutions using larger-scale distributed high end computing platforms including high performance computing for handling the complexity associated with the collaborative solutions for such systems. The study further combined MAS and HPC concepts to overcome the demanding computational requirements of MAS Platforms when used in very large-scale systems. The authors could not include mechanisms to parallelize MAS solutions and repast HPC to solve complex problems in large-scale systems.

Chakraborty & Gupta (2014), proposed medical application using multi-agent system. The study projected on the involvement of multi-agent system in medical or health care domain to solve various medical diagnosis and related problems. Furthermore, a review of different types of approach of multi-agent system in medical domain was presented. The proposed model did not include architectural design for better understanding and clarification.

Chernyshow (2015), designed a multi-Agent system with modeling and identification problems. The study developed a system involving agents using mathematical modeling approach. The design also includes corresponding methods such as notion as knowledge and linear models, which include specific class of non-linear methods. The author did a good job. But the study lack architectural structure for better understanding and clarification.

Chopra & Singh (2015), designed Architecture for Multi-agent Systems based on commitments. The study discussed a middleware that can compute commitments and guarantee alignments between agents in asynchronous settings. The authors did a good job. But they failed to investigate the use of more powerful languages for more subtle situations.

Fahnrich et al, (2015), designed a multi-agent based simulation of decentralized Energy Systems. The authors stated that future energy systems require decentralized regulatory approaches. They developed a simulation model to explore and evaluate appropriate control mechanisms. Although, the described simulation approach integrates the required communication systems and various economic stakeholder groups but could not compare potential migration paths, regulatory incentives and economic impact in other scenarios.

Acharya & Nath (2015), proposed applications of multi-agent systems in control engineering. The study presented a comprehensive survey of the various areas in control engineering where multi-agent systems have been successfully and efficiently implemented. The authors did a good job but failed to identify crucial issues in regard to MAS design and development.

Khalil et al, (2016), looked at multi-agent crisis response systems for the design requirements and analysis of current systems. The study discussed the crisis response domain requirements and provided analysis fo five crisis response namely: DrillSim, DEFACTO, ALADDIN, RoboCup Rescue and FireGride. The authors did a good job. They failed to include the design of self-defensible and adaptable system model for crisis response.

Kumar & Porwal (2015), designed agent-based software testing for multi-agent systems. The study presented the software infrastructure introduced into the agent service framework in order to support ontology design, implementation and management. The authors did a good job. But they could not include ontologies to facilitate interaction among software agents in communication and message validation.

Jemal et al, (2015), designed multi-agent system for hospital organization. The study is an integration of information technology in the health delivery process. The authors could not achieve the following:

1. Introduction of several improvements or extensions to the architecture in agents communication to surmount the problem of semantic operability.

2. Inability to implement a medical ontology with a cloud simulation.

Peter & Wechuli (2016), presented Multi-Agent based surveillance system for Diseases. The study harnesses the use of multi-agent technology based surveillance system for diseases as a solution to the surveillance problems facing the healthcare sector in Kenya and the entire continent of Africa. The study further used an agent oriented methodology called Prometheus for the analysis, design, implementation and evaluation of a program designed to assist medical care practitioners. The authors did a good job. But they failed to carry out a comparative evaluation of their proposed methodology with other existing methodology

Sperka (2016), proposed an overview of multi-agent systems for business process management. The study focused on describing the key concepts of the agent-based system design and traditional business process management approaches. The author did a good job. But there was no architectural model for the design system.

Ciortea et al, (2016), looked at a multi-agent system for manufacturing systems optimization. The study presented a dynamic approach to optimize manufacturing systems based on multi-agent systems. The proposed model also facilitates communication between agents who based on their mechanisms drive to autonomous decision making. The authors did a good job. But could not handle further optimization and flexibility issues in modeling and control.

Okoye et al, (2017), demonstrated a multi-agent based software engineering models. The study critically examined nine different software models for molding and developing multi-agent based systems. The study also showed their various strength and weaknesses in terms of modeling and developing multi-agent based systems. Although the authors did a good job but could not implement the development of any multi-agent based system using the recommended agentology methodology.

Xie & Liu (2017), looked at multi-agent systems and their applications. The study discussed multi-agent systems (MASs) and applications. Furthermore, literature survey on the system architecture, consensus algorithm, multi-agent platform, framework and simulator was also conducted. The authors did a good job. But they could not include ways MASs can be used for commercial purpose. The study also lack technical support for documentation and friendly user interface.

Ciortea et al, (2018), designed repurposing manufacturing lines on the fly with multi-agent systems for the Web-of-Thing (WoT). The study implemented a prototypical production cell that uses industry-grade robots and an augmented reality interface for human workers. The authors did a good job. But they could not envisage a global collective intelligence for manufacturing for the fourth industrial revolution.

Leusin et al, (2018), designed potential of a multi-agent system approach for production control in Smart factories. The study applied a MAS approach to control the production in job shop manufacturing systems to evaluate their performance using a simulation model. However, the study achieves an improved performance with regards to key performance indicators such as work in progress and machine utilization. The study only exploits the possibilities offered by the new technological developments in the context of industry 4.0.

Allamanis & Sutton (2013), presented Mining Source Code Repositories at Massive Scale using Language Modeling and Multi-agent based system. They built the first giga-token probabilistic language model of source code, based on 352 million lines of Java, which they argued that it is 100 times the scale of the pioneering work. Their approach provided a new approach for analyzing software projects, enabling new complexity metrics based on statistical analysis of large corpora, which called data-driven complexity metrics. They proposed a new metrics that measure the complexity of a code module and the topical centrality of a module to a software project. They further explored how the most difficult class of tokens (identifiers) can affect the training procedure and quantified the effect of three types of identifiers. Their result showed that identifiers seem to have a significant role when mining codes. They developed a model that allows token to be successfully predicted across different project domains. Their experiments found that using a large corpus for training these models on code can increase their predictive capabilities. They concluded that the various metrics presented can implemented in large codebases to assist code reuse, evaluate code complexity and assist code navigation and code base visualization. Although they developed a n-gram giga-token model but they could not combine it with other probabilistic LMs models to achieve better predictions.

Chatley & Jones (2018), presented diggit: automated code review via software repository mining for health data. They developed the system as a tool that automatically generate code review comments, offering design guidance on prospective changes, based on insights gained from mining historical changes in source code repositories. They showed that tooling can support and improve code review by automatically extracting relevant information about historical changes and trends from the version control system, and presenting these as part of the review. Table 2.1 shows the summary of related work

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2.1: Summary of related work | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 1. | Abraham et al (2019) | a Machine Learning Approach for Smart Livestock IoT | The authors analyzed various existing supervised and unsupervised machine learning techniques applied in agricultural domain and compare one technique with another with respects to accuracy and a confusion matrix is plotted for each. | The authors did a good job but could not implement the discussed review to a model to show more clarification and understanding. |
| 2. | Fontios et al (2019) | A Review of Machine Learning and IoT in Smart Transportation. | The authors carried out self-contained review of ML (Machine Learning) techniques and IoT applications in Intelligent Transportation Systems (ITS) | The authors did a good job. But the study could not be implemented with real-life health robotics. |
| 3. | Park et al (2019) | Clot-Net: A Scalable Cognitive IoT based Smart City Network and Architecture. | The authors proposed a CIoT-based smart city network (CIoT-Net) architecture which describes how data gathered from smart city applications can be analyzed using cognitive computing and handle the scalability and flexibility problems. | The authors did a good job. However, a major limitation of their study is that the developed model was deficient in benchmarking and cost benefits analysis. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 4. | Stefano *et al* (2015) | An agent-based architecture for adaptive supervision and control of smart environments. | The study described architecture and functionality of a generic agent that is in charge of handling a given environment in an Ambient Intelligence context, ensuring suitable contextualized and personalized support to the user’s actions, adaptively to the user’s peculiarities and to changes over time. | The authors did a good job. However, the analysis of their adopted methodology showed that they only simulated the implementation, and failed to deploy the work to a real smart environment. |
| 5. | Juan *et al* (2010) | An agent-based architecture for developing activity aware systems for assisting the elderly. | The authors proposed an activity-aware computing that allows smart environments to provide continuous activity awareness | The authors did a good job. However, they were unable to apply the developed system on other software engineering tasks that rely on text analysis using topic models. |
| 6. | Pekka *et al* (2017) | In-home Advanced Robotic System . | Examined the safety profile and usability of an integrated advanced robotic device and telecare system to promote medication adherence for elderly home-care patients. | However, their developed model failed to proffer solution to identified cases of missed doses. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 7. | Ayman *et al* (2019) | Adaptive intelligent alarm system for wireless sensor network. | The study proposed a basic and adaptable remote arrange for domestics computerization of temperature, moistness, gas, movement and light by executing dependable sensor hubs which can be controlled too observed. | However, they had a vague result due to their simulation and non-implementation with a real hardware sensor device. |
| 8. | Debajyoti *et al* (2018) | The elderly users’ adoption of smart home services | The study proposed and validated a new comprehensive research model called the elderly smart home technology acceptance model by extending the original technology | However, the analysis of their adopted methodology showed that they only simulated the implementation, and failed to deploy the work to a real smart home area. |
| 9. | Ashalatha *et al* (2012) | Architecture modeling and formal analysis of intelligent multi-agent systems. | According to the study, modern cyber-physical systems usually assume a certain degree of autonomy. Such systems, like Ambient Assisted Living systems aimed at assisting elderly people in their daily life. | However, the datasets used for training and designing their model was not a hybrid of supervised and unsupervised learning techniques which further resulted to latencies in their model performance. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 10. | Zaki and Abdus (2017) | Machine learning in wireless sensor network. | Within the study, a concept of machine learning strategies was suggested. Furthermore, the authors carried out investigation to address design issues in wireless sensor networks. The authors did a good job. | However, the suggested strategies were not tested with a machine-learning oriented model. |
| 11. | Aderemi *et al* (2016) | Development of smart assistive DTMF home automated system for ageing population. | The authors designed and implemented a cost-effective smart assistive DTMF home automation system that utilizes a tele-remote circuit to control home appliances via existing cellular communication networks. | The authors did a good job. However, the datasets used for training and designing their model was not a hybrid of supervised and unsupervised learning techniques which further resulted to latencies in their model performance. |
| 12. | Bethany *et al* (2017) | Evolution of smart homes for the elderly | The study presented existing technologies in the evolution of smart homes for the elderly where development is lacking. | However, the suggested smart technologies if implemented will lack cost-benefit analysis for the mentioned elderlies. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 13. | Pavlos and Nikolaos (2011) | Combining Gaia and JADE for multi-agent systems development. | The authors presented an enhanced version of a roadmap they had previously proposed, concerning how one can implement JADE agents using the Gaia methodology for analysis and design purposes. | The authors did a good job but failed to implement the discussed roadmaps to a machine-learning oriented model for more clarification and understanding. |
| 14. | Jamie *et al* (2017) | Healthcare in the Smart Home: a study of past, present and future. | The study looked at the history of Smart Home Healthcare, current research areas, and potential areas of future investigation. | The authors did a good job. However, the reviewed history on healthcare was not practically implemented to a model. |
| 15. | Marijke *et al* (2018) | Activities of daily living in older community dwelling persons. | The authors analyzed Electronic databases (Medline, EMBASE, AMED, Psycinfo, CINAHL), using MeSH terms and relevant keywords. Studies, published in English, were included if they evaluated one or more psychometric properties of ADL instruments in community-dwelling | The authors did a good job. However, they failed to utilize a biometrical approach to the integration of health-related electronic databases. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 16. | Basma *et al* (2014) | Independent Living for Persons with Disabilities and Elderly People using Smart Home Technology. | The study took a snapshot on the state of the art in the smart home technology for elders and disabilities people. Also it proposed a new Elders/Disabilities Wireless Smart Home for assistive independent living (E/D-WSH) with its approximate cost compared with the Indian home automation market. | The proposed system was not backed up with real-time data hardware devices for the discussed smart home technology. |
| 17. | Ismail and Erkan (2017) | A smart house application using wireless sensor networks. | In the study, a smart home via smart phones and pcs was explained. In addition, the control of electrical and electronic devices in houses and buildings was also scrutinized by the authors. | The authors did a good job. However, performance evaluation of their developed model showed some routing deficiencies. |
| 18. | Toshifumi (2015) | In-home Health monitoring system for solitary elderly. | In the study, the authors proposed a sensor-based monitoring system that evaluates the health status of solitary elderly based on daily living activities, and provides forecasts of emergency situations | The authors did a good job. However, the implementation carried out by them was only limited to few trials. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 19. | Heetae *et al* (2018) | IoT Smart Home Adoption | The study examined the smart home service features that current users require and empirically evaluates the relationship between the critical factors and the adoption behavior with 216 samples from Korea. | However, the obtained result from the study was deficient in cost benefit analysis. |
| 20. | Salman *et al* (2011) | Real Time Algorithm for the Smart Home Automation based on the Internet of Things. | The authors proposed an algorithm for smart home automation system based on IoT using sensor nodes which are directly connected to Arduino Nano. | The authors did a good job but could not further implement with IoT using sensor nodes which are directly connected to Arduino Nano |
| 21. | Dagmawi *et al* (2019) | Reasoning in Multi-agent based smart phones | The study examined literature on home systems and provides a comprehensive overview of the essential requirements, assumptions, strengths, limitations, challenges and future research directions of their proposed reasoning systems. | However, the discussed findings from the study were not implemented with a machine learning oriented model. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 22. | Akshaya *et al* (2018) | Smart Pill Box. | The study proposed a mobile application called smart pill box which addresses medical issues of the elderly. Smart pill box alerts the patient by giving an alarm signal to take the medicines in the correct time. | However, comparative analysis of the Smart Pill Box shows that every user must possess a smart phone without which they cannot get pill notifications and communicate with pharmacy or their caretaker. |
| 23. | Sumit *et al* (2017) | Smart Homes for Elderly Healthcare. | The study presented a comprehensive review on the state-of-the-art research and development in smart home based remote healthcare technologies. | The reviewed study on smart homes for the elderly healthcare was not implemented to a model for more clarification and understanding. |
| 24. | Quynh *et al* (2012) | Smart Homes for Older People | The study examined the concept of smart homes in a technologically driven society. Furthermore, the study discussion was focused on challenges in the use of smart homes among older people such as accessibility and ethical issues. | The authors did a good job. However, their proposed system was not backed up with real-time data hardware devices for the discussed smart home technology. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 25. | Rameshkumar and Viswanathan (2019); | Multi-agent based architecture design for wireless Body Area Network Monitoring System. | In their study, they developed multi-agent system architecture for Patient Monitoring (PM) which leads to the provision of reliable and cheap health care for the elderly. | However, they failed to implement a firewall for embedded security and trust during transmission of health information via sensors. |
| 26. | Achim *et al* (2019) | A decision support system in back pain diagnosis. | The study investigated the concordance of a decision support system and the recommendation of spinal surgeons regarding back pain. Furthermore, eleven (11) patients completed the decision support system in which their illness was diagnosed by a spinal surgeon. | The results of the study did not show significant medium relation between the decision support system and diagnosis of the medical doctor. |
| 27. | Mohammed *et al* (2011) | Design and implementation of fuzzy expert system for back pain diagnosis. | The authors produced a Fuzzy Expert System (FES) to diagnosis of back pain disease based on the clinical observation symptoms using fuzzy rules. | There was no adequate comparative analysis between their Fuzzy Expert System and other expert systems. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 28. | Ekong (2013) | Fuzzy Inference System for Predicting Depression Risk Levels | The study described research results in the development of a fuzzy driven system to determine the depression risk levels of patients. The system was implemented and simulated using MATLAB fuzzy tool box. | The author did a good job but failed to illustrate the drawbacks of software intelligence that could be capitalized by hackers. |
| 29. | Sabreen and Naser (2017) | Expert system for diagnosing ankle diseases | The authors developed an expert system that is based on the principle of asking the user gradual questions about the symptoms he feels, leading him to the result of diagnosing the illness, dealing with it quickly, and tips for permanent treatment. | These systems may also help both trainee physicians and physiotherapists. There were identified latencies during performance evaluation of the system. |
| 30. | Matthias *et al* (2017) | Expert System for Medical Diagnosis of Hypertension and Anemia. | The study implemented a model for diagnosing Hypertension and Anemia using Bayesian Network technology which is efficient for modeling real-life problems and reasoning under certainties. | However, their utilized SQL Database was only limited to structured datasets only. |
| Table 2.3: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 31. | Baki *et al* (2018) | CASE: a framework for computer supported outbreak detection. | The authors described the design and implementation of a computer supported outbreak detection system called CASE (named after the protagonist of the William Gibson novel Neuromancer), or Computer Assisted Search for Epidemics. | However, the author failed to implement their system with a secured embedded IoT device. |
| 32. | Madhulatha (2012) | An overview on clustering methods for disease detection | The study covered about clustering algorithms, benefits and its applications. According to the author, clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis | However, analysis of the study showed that the author failed to implement the discussed clustering methods to a model for further clarification and understanding. |
| 33. | Zhiting *et al* (2018) | Unification of deep generative models | According to the authors, “deep generative models have achieved impressive success in recent years. Generative Adversarial Networks (GANs) . | The authors failed to adopt a supervised learning approach to their designed paradigm for deep generative models. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 34. | Jungang *et al* (2015) | Overview of deep generative models. | The study analyzed three important deep generative models including DBNs, deep autoencoder, and deep Boltzmann machine are reviewed. In addition, some successful applications of deep generative models in image processing, speech recognition and information retrieval are also introduced and analyzed. | However, analysis of the study showed that the authors failed to adopt a supervised learning approach to their designed paradigm for deep generative models. |
| 35. | David and Graham (2012) | Design science in decision support systems research | According to the authors, “design science has been an important strategy in decision support systems (DSS) research since the field’s inception in the early 1970s”. | The author failed to implement the discussed decision support system methods to a model for further clarification and understanding. |
| 36. | Smita and Patel (2016) | Study of graph storage database of NoSQL for Health Information | The study described what big data storage management is, dimensions of big data, types of data, what is structured and unstructured data. | However, the authors failed to implement the discussed big data concepts with NoSQL database. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 37. | Wasiwasi and Zaipuna (2014) | Design and development of web-based digital repository for elderly’s communication. | The aim of the study was to design and develop a web‐based digital repository for elderly communications using NM‐AIST as a case study. The system was developed using open source software. | The authors did a good job. However, a major limitation of their study is that the developed model was deficient in benchmarking and cost benefits analysis. |
| 38. | Ieva *et al* (2019) | Deploying Smart Program Understanding on a Large Code Base. | They argued that new developers face the task of classifying a huge number of files and functions without any help. Their proposed approach to this problem called FEAT automatically extracts topoi from source code by using hierarchical agglomerative clustering. | But they could not use semantic classification of all projects source codes in the automatic detection of semantic clones at a very large scale as well as mining of test script to find hidden links between test cases. |
| 39. | Ieva *et al* (2018) | Discovering Program Topoi through Clustering. | They stated that organizing the code in a more abstract way, closer to humans is an attempt that received interest from the software engineering community | But unfortunately, there is no recognized recipe or tool that can concretely provide any help in dealing with large software repositories. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 40. | Allamanis and Sutton (2013) | Mining Source Code Repositories at Massive Scale using Language Modeling and Multi-agent based system. | They built the first giga-token probabilistic language model of source code, based on 352 million lines of Java, which they argued that it is 100 times the scale of the pioneering work. | Although they developed a n-gram giga-token model but they could not combine it with other probabilistic LMs models to achieve better predictions. |
| 41. | Chatley and Jones (2018) | Diggit: Automated Code Review via Software Repository Mining for Health Data. | The developed the system as a tool that automatically generate code review comments, offering design guidance on prospective changes, based on insights gained from mining historical changes in source code repositories. | Although the developed system could automatically take action on code review, it only approximates developer’s action on analysis suggestions. |
| 42. | Thomas (2010) | Mining Software Repositories for Health Data Management using Topic Models. | The author proposed the use of statistical topic models to automatically discover structure in these textual repositories stating that Software repositories such as source code repository contain data that are unstructured, unlabeled | However, a major limitation of their study is that the developed model was deficient in benchmarking and cost benefits analysis. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 43. | Rangaswamy and Shobha (2012) | Optimized Association Rule Mining Using Genetic Algorithm. | They proposed a system; Optimized Association Rule Mining (OARM) that applies genetic algorithm over the rules fetched from apriori association rule mining. | But they could not improve the client/server interaction in the system by minimize the complexity of the genetic algorithm and scanning of database. |
| 44. | Panichella *et al* (2013) | How to effectively use topic models for Software Engineering Tasks? An Approach Based on Genetic Algorithms. | They proposed a novel solution to adapt, configure and effectively use a topic modeling technique, namely Latent Dirichlet Allocation (LDA), to achieve better (acceptable) performance across various Software Engineering tasks. | But they were unable to apply the developed system on other software engineering tasks that rely on text analysis using topic models. |
| 45. | Munaiah *et al* (2017) | Curating GitHub for Engineered Software Projects | They presented a reference implementation of the framework as a tool referred to as Reaper, to enable researchers to select GitHub repositories that contain evidence of an engineered software project. | But there is a limitation in the system; it does not collect dimensions metric from a repository, which can induce bias in a selected repository. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 46. | Bradley and Murphy (2011) | Supporting Software History Exploration for Home Management of the Elderlies. | They proposed a user interface, in a tool called Rationalizer that integrates historical information into the source code editor, supporting exploration from a particular code line to its immediate history. | However, the system could not determine the relative frequency in real-world development of the types of historical questions was identified. |
| 47. | Upadhyaya and Rajan (2018) | Accelerating Source Code Analysis of Health Data at Massive Scale. | They proposed a technique that reduces the amount of computation performed by the ultra-large-scale source code mining task, especially those that make use of control and data flow analyses. | But they could not investigate whether the similarity discovered in nodes and edges of the reduced control flow graph (RCFG) can be used to run the mining task on only unique reduced control flow graphs to yield better results. |
| 48. | Ray *et al* (2017) | Large-Scale Study of Programming Languages and Code Quality in Github for Health Information Management. | They presented a large-scale study of language type and use as it relates to software quality using data extracted from Github characterized by its complexity and variation along multiple dimensions. | However, they were unable to carry out additional survey to quantify the specific effects of language type on usage. |
| Table 2.1: Summary of related work (Contd.) | | | | |
| **SN.** | **AUTHOR(s)** | **RESEARCH CARRIED OUT** | **SUMMARY OF FINDINGS** | **LIMITATIONS** |
| 49. | Todor and Juri (2017) | Expert System for Milk and Animal Monitoring. | The study presented design and development of an expert system for milk and animal monitoring. | The developed model in their study failed to monitor other aspect of livestock especially animal husbandry in order to detect any form of inhumanity to animals by a farm worker. |
| 50. | Andzio and Xiong (2019), | Agriculture Monitoring System using Smart and Innovative Farming. | The authors developed an optimal crop irrigation/agriculture system based on a network of wireless sensors. The work intended to strategize and maintain a control organism utilizing crop sensor that senses with data managing with a web application and a smartphone. | However, the study direction was identified to focus mainly on irrigation monitoring and fisheries. It failed to monitor other aspect of livestock especially animal husbandry in order to detect any form of inhumanity to animals by a farm worker. |

**2.6 Direction of the study**

Having reviewed several related literature as summarized in Table 2.1, the study showed keen interest in the most current and closely related issue. Todor and Juri (2017), looked at Expert System for cattle milking and monitoring. The study presented the design and development of an expert system for data collection, analysis and decision of farm cattle management. Furthermore, the authors also opined that Expert systems (ES) are characterized with high performance, high responsiveness and reliability combined with ease of use and ease of understanding, which makes developing of such system challenging and involves different type of specialists. However, the developed model in their study failed to implement a fuzzy-based model for real-time monitoring of smart livestock farming. This would have facilitated real-time monitoring of the farmer’s behavior to the cattle and further report any form of inhumanity to the cattle by the farmer. This study intends to enhance the work of Todor and Juri (2017) with a fuzzy logic model for monitoring smart livestock farming.

**CHAPTER THREE**

**MATERIALS AND METHODS**

**3.1 Analysis of the Existing System**

The existing system is an existing system is an expert system for milk and animal monitoring as illustrated in figure 3.1. The system also presented design and development of an expert system for data collection, analysis and decision using decision tree algorithm. The animal monitoring system is planned as expert system for the dairy industry. It is mainly targeted for cows but with adaptation of the algorithm should be able to work with goats and buffalos. It focuses on both milk quality and animals health. It tries to detect early symptoms of mastitis, based on specific physical qualities of the milk samples and start early treatment of the animal and early separation of infected milk, preventing it from mixing with other milk in the tanks which can potentially lead to infecting the whole tank with pathogens and cause significant loses for the farmers from the order of thousands of euro depending on the tank size.

The authors of the existing system model also opined that the use of efficient procedures and rules by the inference engine is essential in deducting a correct and flawless solution. In case of knowledge based expert system, the inference engine acquires and manipulates the data in the knowledge base to arrive in particular solution. In case of rule based expert system, it applies rules repeatedly to the facts, which are obtained from earlier rule execution. It might also add new knowledge to the knowledge database if required and resolve rule conflicts when multiple rules are applicable to a particular case. In order to recommend a solution, the engine might use forward chaining or backward chaining strategy. The Inference engine runs the chain of conditions and derivation in order to deduce the outcome.

**USER 1**

**USER 2**

**USER N**

**Cattle Milking and Monitoring**

**Cattle Monitoring Reports**

**Cattle Monitoring Reports Output**

**Knowledge-Base**

**Inference Engine**

**User Interface**

Figure 3.1: Existing System Architecture for Cattle Monitoring

(**Source:** Todor and Juri, 2017)

**3.1.1 Detailed Analysis of the Existing System**

This section discusses each component of the Existing System Architecture

**SPECIFICATIONS**

**i) Coding:**

Python,

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**User**

Figure 3.2: End-user component of the Existing System Architecture

This component represents a person who ultimately uses the system. The end user also stands in contrast to users who support or maintain the system. In addition, the terminal for the end-user usually consist of a keyboard and monitor, that acts as a front end for a mainframe, terminal server, or other back-end processing device.

**SPECIFICATIONS**

**i) Coding:**

Python,

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Knowledge-Base**

Figure 3.3: Knowledge-Base component of the Existing System Architecture

This component illustrates a published collection of documentation that typically includes answers to frequently asked questions, how-to guides, and troubleshooting instructions. Its purpose is to make it easy for people to find solutions to their problems without having to ask for help. As expert systems moved from being prototypes to systems deployed in corporate environments the requirements for their data storage rapidly started to overlap with the standard database requirements for multiple, distributed users with support for transactions.

**SPECIFICATIONS**

**i) Coding:**

Hypertext Pre-processor

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Inference Engine**

Figure 3.4: Inference Engine component of the Existing System Architecture

This component illustrates a tool used to make logical deductions about the knowledge-base. Experts often talk about the inference engine as a component of a knowledge base. Inference engines are useful in working with all sorts of information, for example, to enhance business intelligence. An inference engine is often a component of a knowledge base combined with the knowledge base; the inference engine helps stakeholders to get those logical insights from the storehouse of information at their disposal

**User Interface**

**SPECIFICATIONS**

**i) Coding:**

Laravel

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

User Interface

Username

Password

Test-Set Input

Figure 3.5: User Interface Component of the Existing System Architecture

This component illustrates the access points where users interact with designs. They come in three formats: Graphical user interfaces (GUIs). Users interact with visual representations on digital control panels. A computer's desktop is a GUI. In addition, the user interface (UI) is the space where interactions between humans and machines occur. The goal of this interaction is to allow effective operation and control of the machine from the human end, whilst the machine simultaneously feeds back information that aids the operators' decision-making process.

**SPECIFICATIONS**

**i) Coding:**

Java

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Cattle milking and monitoring**

**Cattle monitoring reports**

**Cattle monitoring reports output**

Figure 3.6: Farm Monitoring Component of the Existing System

This component illustrates the monitoring processes of the farm activities, especially for that of cattle. It also consists of sub-components for farm monitoring such as the milk and animal monitoring, analysis of animal monitoring reports and outputted monitoring results. In addition, the component makes it possible to monitor the farm constantly, measuring data such as smoke, flooding, air temperature and humidity, the temperature of liquids and power cuts originating from cellars, barns, stables, silos, residential buildings and greenhouses.

**3.1.2 Dataflow diagram of the Existing System**

The dataflow diagram (DFD) is a way of representing a flow of data through a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself (Figure 3.7). A data-flow diagram has no control flow; there are no decision rules and no loops. The DFD consists of processes, flows, warehouses, and terminators. There are several ways to view these DFD components.

The process (function, transformation) is part of a system that transforms inputs to outputs. The symbol of a process is a circle, an oval, a rectangle or a rectangle with rounded corners (according to the type of notation). The process is named in one word, a short sentence, or a phrase that is clearly to express its essence.

Data flow (flow, dataflow) shows the transfer of information (sometimes also material) from one part of the system to another. The symbol of the flow is the arrow. The flow should have a name that determines what information (or what material) is being moved. Exceptions are flows where it is clear what information is transferred through the entities that are linked to these flows. Material shifts are modeled in systems that are not merely informative. Flow should only transmit one type of information (material). The arrow shows the flow direction (it can also be bi-directional if the information to/from the entity is logically dependent - e.g. question and answer). Flows link processes, warehouses and terminators.

In addition, the DFD is a system created by analysts based on interviews with system users. It is determined for system developers, on one hand, project contractor on the other, so the entity names should be adapted for model domain or amateur users or professionals. Entity names should be general (independent, e.g. specific individuals carrying out the activity).

Initialization of the System

Request to view Farm Monitoring Reports

Authentication of Registered Users

Input (Xi)

Results

New User Registration

The system is set-up and connected

User input are scanned, matched and authenticated on the System

Display of Farm Monitoring Reports

Key input of user details

Downloading Farm Monitoring Reports

Figure 3.7: Dataflow Diagram of the Existing System

(Source: Todor and Juri, 2017)

**3.1.3 Decision tree algorithm of the Existing System**

The following is the decision tree algorithm of the existing system:

INPUT: S, where S = set of classified instances

OUTPUT: Decision Tree

Step 1: Start

Step 2: Initialize System for Milk and Animal Monitoring

Step 3: Procedure BUILD TREE

Step 4: repeat

Step 5: MaxGain 0

Step 6: SplitA null

Step 7: e Entropy (Attributes)

Step 8: For all Attributes a in S do

Step 9: gain informationGain (a,e)

Step 10: if Gain > maxGain then

Step 11: MaxGain gain

Step 12: Split A a

Step 13: end if

Step 14: Partition (S, Split A)

Step 15: end procedure

**3.1.3 Advantages of the Existing System**

The following are advantages of the existing system:

1. ability to monitor livestock activities especially the environment, animal and other farm processes

**3.1.4 Disadvantages of the Existing System**

The existing system failed to implement a fuzzy-based model for real-time monitoring of smart livestock farming. This would have facilitated real-time monitoring of the farmer’s behavior to the cattle and further report any form of inhumanity to the cattle by the farmer.

**3.2 Design of the Proposed System**

The new system is the enhancement of the existing system model. The new enhancement encompasses an improved smart model for livestock monitoring as shown in figure 3.8. The smart model will encompass a rule-base and framework for animal monitoring and interpretation. The rule-base will consist of a knowledge-base and inference engine which will store and manipulate knowledge to interpret livestock activities in a visualized and useful way. Furthermore, the framework for data visualization will also optimize the system performance via visual elements such as charts, graphs, and maps, in order to provide an accessible way to see and understand trends, outliers, and patterns in data.

In addition, the concept of the new model is to ensure that employee farmers in large-scale farms are nicer to the farm animals so as to optimize farm productivity.

**USER 1**

**USER 2**

**USER N**

**Cattle milking and monitoring**

**Cattle monitoring reports**

**Cattle monitoring report outputs**

**Knowledge-Base**

**Inference Engine**

**User Interface**

**FUZZY LOGIC MODEL**

**Visualized monitoring from embedded sensor in farmer’s clothes**

**Fuzzy logic technique**

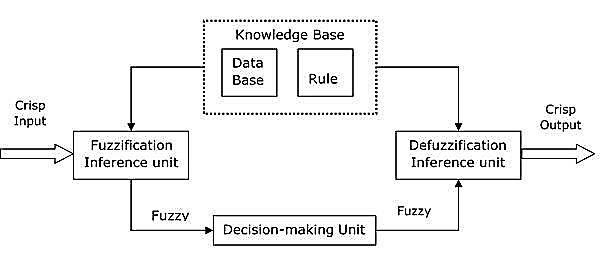


Figure 3.8: Architectural design of the Proposed System

**3.2.1 Detailed analysis of the Proposed System**

This section discusses each component of the proposed system architecture

**SPECIFICATIONS**

**i) Coding:**

Python,

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**User**

Figure 3.9: End-user component of the Proposed System Architecture

This component represents a person who ultimately uses the system. The end user also stands in contrast to users who support or maintain the system. In addition, the terminal for the end-user usually consist of a keyboard and monitor, that acts as a front end for a mainframe, terminal server, or other back-end processing device.

**SPECIFICATIONS**

**i) Coding:**

Python,

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Knowledge-Base**

Figure 3.10: Knowledge-Base component of the Proposed System Architecture

This component illustrates a published collection of documentation that typically includes answers to frequently asked questions, how-to guides, and troubleshooting instructions. Its purpose is to make it easy for people to find solutions to their problems without having to ask for help. As expert systems moved from being prototypes to systems deployed in corporate environments the requirements for their data storage rapidly started to overlap with the standard database requirements for multiple, distributed users with support for transactions.

**SPECIFICATIONS**

**i) Coding:**

Hypertext Pre-processor

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Inference Engine**

Figure 3.11: Inference Engine component of the Proposed System Architecture

This component illustrates a tool used to make logical deductions about the knowledge-base. Experts often talk about the inference engine as a component of a knowledge base. Inference engines are useful in working with all sorts of information, for example, to enhance business intelligence. An inference engine is often a component of a knowledge base combined with the knowledge base; the inference engine helps stakeholders to get those logical insights from the storehouse of information at their disposal

**User Interface**

**SPECIFICATIONS**

**i) Coding:**

Laravel

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

User Interface

Username

Password

Test-Set Input

Figure 3.12: User Interface Component of the Proposed System Architecture

This component illustrates the access points where users interact with designs. They come in three formats: Graphical user interfaces (GUIs). Users interact with visual representations on digital control panels. A computer's desktop is a GUI. In addition, the user interface (UI) is the space where interactions between humans and machines occur. The goal of this interaction is to allow effective operation and control of the machine from the human end, whilst the machine simultaneously feeds back information that aids the operators' decision-making process.

**SPECIFICATIONS**

**i) Coding:**

Java

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Cattle milking and monitoring**

**Cattle monitoring reports**

**Cattle monitoring reports outputs**

Figure 3.13: Farm Monitoring Component of the Proposed System

This component illustrates the monitoring processes of the farm activities, especially for that of cattle. It also consists of sub-components for farm monitoring such as the milk and animal monitoring, analysis of animal monitoring reports and outputted monitoring results. In addition, the component makes it possible to monitor the farm constantly, measuring data such as smoke, flooding, air temperature and humidity, the temperature of liquids and power cuts originating from cellars, barns, stables, silos, residential buildings and greenhouses.

**SPECIFICATIONS**

**i) Coding:**

Python

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

**Visualized monitoring from embedded sensor in farmer’s clothes**

3.14: Visualized Monitoring Component of the Proposed System

This component illustrates the real-time monitoring activities of the farm cattle. Its method of communication encompasses online animated video of the farmer’s management activities. This is also called data visualization which is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. It is also an interdisciplinary field that deals with the graphic representation of data. It is a particularly efficient way of communicating when the data is numerous as for example a time series

**Fuzzy logic technique**

**SPECIFICATIONS**

**i) Coding:**

Java

MySQL

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

Figure 3.15: Fuzzy logic component of the Proposed System Architecture

Fuzzy logic is a branch of machine learning that is being utilized to determine partial truth of a computing process using Boolean value which includes 0 and 1. Fuzzy logic algorithm helps to solve a problem after considering all available data. Then it takes the best possible decision for the given the input. It is a concept that is still being applied in industrial areas such as manufacturing, electricity, robotics, etc.

**Rule-Base**

**SPECIFICATIONS**

**i) Coding:**

Python,

Java

R

Apache Cassandra

BOOTSTRAP

ii) **Scripting:**

JavaScript

iii)**Type:** Simulator

Figure 3.16: Rule-Base Component of the Proposed System Architecture

This component is used to store and manipulate knowledge to interpret information in a useful way. It is often used in artificial intelligence applications and research.

Normally, the term rule-based system is applied to systems involving human-crafted or curated rule sets. Rule-based systems constructed using automatic rule inference, such as rule-based machine learning, are normally excluded from this system type.

**3.2.2 Dataflow diagram of the Proposed System**

The dataflow diagram (DFD) is a way of representing a flow of data through a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself (Figure 3.17). A data-flow diagram has no control flow; there are no decision rules and no loops. The DFD consists of processes, flows, warehouses, and terminators. There are several ways to view these DFD components.

The process (function, transformation) is part of a system that transforms inputs to outputs. The symbol of a process is a circle, an oval, a rectangle or a rectangle with rounded corners (according to the type of notation). The process is named in one word, a short sentence, or a phrase that is clearly to express its essence.

Data flow (flow, dataflow) shows the transfer of information (sometimes also material) from one part of the system to another. The symbol of the flow is the arrow. The flow should have a name that determines what information (or what material) is being moved. Exceptions are flows where it is clear what information is transferred through the entities that are linked to these flows. Material shifts are modeled in systems that are not merely informative. Flow should only transmit one type of information (material). The arrow shows the flow direction (it can also be bi-directional if the information to/from the entity is logically dependent - e.g. question and answer). Flows link processes, warehouses and terminators.

In addition, the DFD is a system created by analysts based on interviews with system users. It is determined for system developers, on one hand, project contractor on the other, so the entity names should be adapted for model domain or amateur users or professionals. Entity names should be general (independent, e.g. specific individuals carrying out the activity).

Initialization of the System

Initialization of the Monitoring Model

Activate Real-time Visualized Monitoring Process using Fuzzy Logic

Authentication of Registered Users

Input (Xi)

Results

New User Registration

The system is set-up and connected

User input are scanned, matched and authenticated on the System

Display of Farm Monitoring Reports

Key input of user details

Real-time Monitoring Process

Figure 3.17: Dataflow Diagram of the Proposed System

**3.2.3 Mathematical modeling of the Proposed System**

The following are equations of the proposed system

**Step 1:**

Let the fuzzy set A ⊂∼ U be given by the membership function for livestock system optimization where L = [0,1]. Let A,B ⊂∼ U. Then A ⊆ B if A(x) ≤ B(x) holds for all x ∈ U. The set of all fuzzy sets on U is

**Step 2:**

F(U) = {A |A ⊂∼U}= LU (3.1)

**Step 3:**

Note that F(U) contains also all ordinary subsets of U.

Given the fuzzy set A ⊂∼ U, it can be characterized in several ways.

**Step 4:**

Supp(A) = {x |A(x) > 0} (3.2)

where Supp means support

**Step 5:**

Aa = {x |A(x) ≥a}, a ∈(0,1] (3.3)

**Step 6:**

Ker(A) = {x |A(x) = 1}. (3.4)

where Ker means kernel

**Step 7:**

The fuzzy set A for livestock system optimization is normal if Ker(A) = ∅. The empty fuzzy set is defined by

∅= {0/x |x ∈U} (3.5)

**3.2.4 Fuzzy Logic algorithm of the Proposed System**

The following is the proposed system algorithm

**Algorithm 2:** Fuzzy Logic system for livestock data optimization

**Procedure:** Fuzzification and Defuzzification of key information from livestock data

**Input:**  Crisp sets

**Output:** Defuzzified information

1. Initialize Fuzzy System
2. Input crisp sets for optimal performance of pre-processed livestock data
3. Fuzzify crisp inputs using knowledge-base
4. x0  x1 x2

5. x = [[1., 0., 0.],

6. [1., 0., 1.],

7. [1., 1., 0.],

8. Display fuzzified crisp datasets as Output

9. End

**3.2.5 Methodology**

The method adopted by the study for the proposed system design is the water-fall model. The Waterfall model is the earliest SDLC (Software Development Life-cycle) approach that was used for software development. The waterfall Model illustrates the software development process in a linear sequential flow.

The sequential phases in Waterfall model are:

Requirement Gathering and analysis − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.

System Design − The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.

Implementation − With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing. Integration and Testing − All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

**3.2.6 Frontend Interface Design of the Proposed System**

Figure 3.6 shows the Frontend Design of the Proposed System

REGISTRATION PAGE

WELCOME PAGE

NETBEANS INITIALIZATION FOR PYTHON

CRISP SET INPUT PAGE

LOGIN PAGE

REAL-TIME VIDEO MONITORING AND VISUALIZATION

Figure 3.18: Frontend Interface Design of the Proposed System

**3.2.7 Steps for the Proposed System**

The proposed system steps were depicted with a Use-Case Activity Diagram as illustrated in 3.19. A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses. It is a written description of how users will perform tasks on your website. It outlines, from a user's point of view, a system's behavior as it responds to a request. Each use case is represented as a sequence of simple steps, beginning with a user's goal and ending when that goal is fulfilled.

A use case diagram is a behavior diagram and visualizes the observable interactions between actors and the system under development. The diagram consists of the system, the related use cases and actors and relates these to each other:

System: What is being described?

Actor: Who is using the system?

Use Case: What are the actors doing?

A use case diagram describes the desired functionality of the system and relates it to use cases and actors. That way it can represent existing viewpoints of the system and how they are interpreted differently only through this can requirements be completely understood. Use cases are normally presented as ovals. A use case represents a functionality of the system from the viewpoint of the user and describes the goals of their use. Here, the order of the action can and should be entered. For example insert card, enter PIN, and so on.

**End-user**

Figure 3.19: Use-Case Activity Diagram of the Proposed System Design

**3.2.8 Unified Modeling Language of the Proposed System**

Figure 3.20 shows the UML diagram of the Proposed System

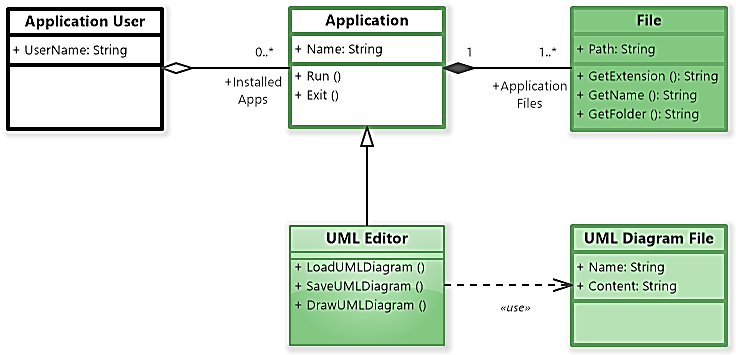


Figure 3.20: UML Diagram of the Proposed System

**3.3 Datasets**

This section discusses the means collecting all the information for training and testing the proposed system and also managing it in a way that maximizes the speed and comprehensiveness with which critical information can be extracted, analyzed and used.

**3.3.1 Training Set**

The training set was adopted during the learning process of the proposed system. Furthermore, a supervised learning algorithm looks at the training dataset to determine, or learn, the optimal combinations of variables that will generate a good predictive model. The goal of the training process is to produce a trained (fitted) model that generalizes well to proposed, unknown data. The fitted model is evaluated using “proposed” examples from the held-out datasets (validation and test datasets) to estimate the model’s accuracy in classifying proposed data.

**3.8.2 Test Set and Method of Data Collection**

The test-set was obtained from dispatched surveys on livestock farming is independent of the training dataset, but follows the same probability distribution as the training dataset. If the proposed system fits to the training set, it also fits the test. Furthermore, a better fitting of the training dataset as opposed to the test dataset usually points to over fitting. In addition, the test-set is used only to assess the performance of the proposed system

Percentage of data used for training (%) = 70

Percentage of data used for testing (%) = 30

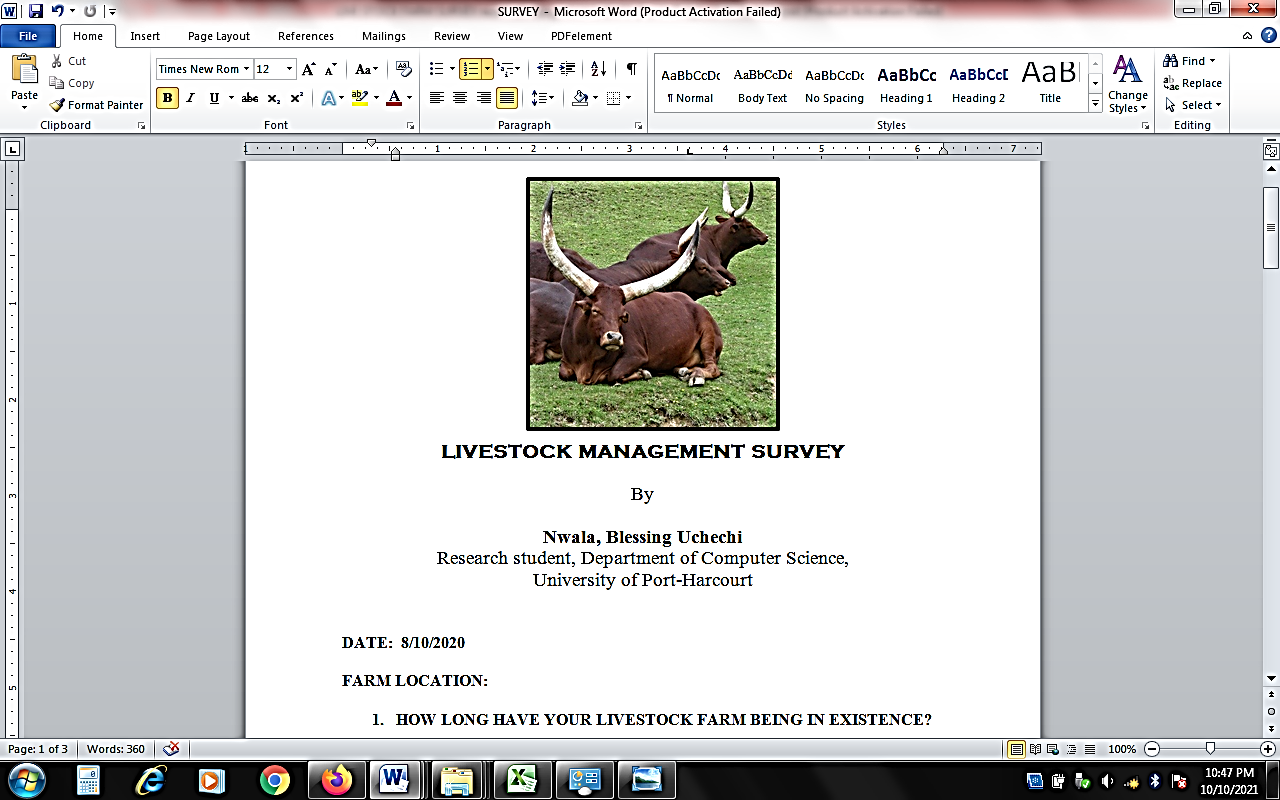


Figure 3.21: Livestock management survey for data collection and analysis

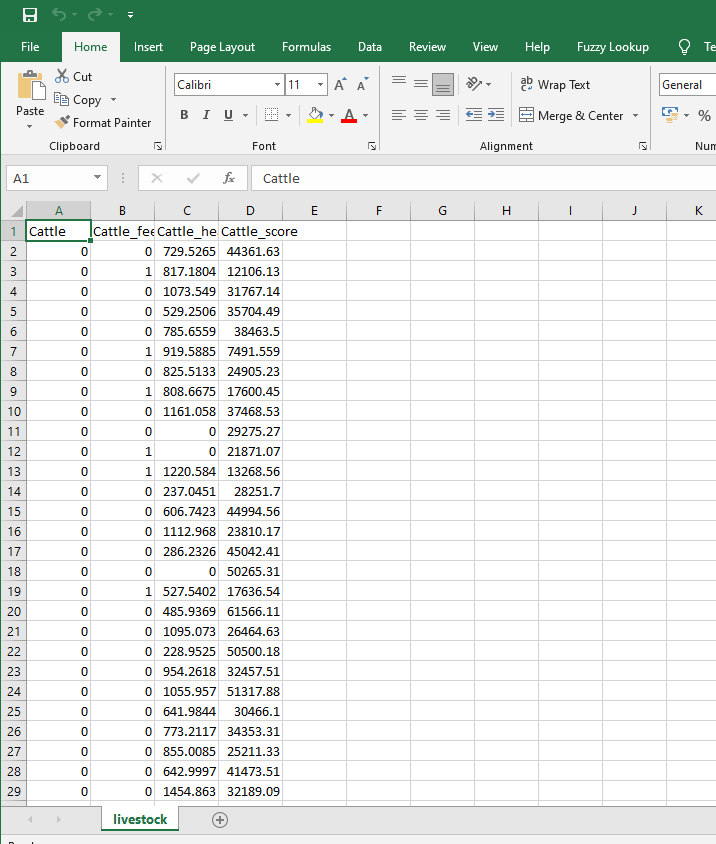


Figure 3.22: Livestock CSV Data set

(**Source:** https://data.ilri.org/portal/dataset?vocab\_ILRI\_vocsubjects=livestock)

Table 3.2: Test-set for the proposed system implementation

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sn** | **Fname** | **Lname** | **Gender** | **Email** | **Address** | **Phone\_No:** | **User**  **name** | **Password** |
| 1. | Mafuzu | Clark | Male | mc@gmail.com | Null | Null | Kun | P01 |
| 2. | Sarah | Williams | Female | sw@gmail.com | Null | Null | Sav | P02 |
| 3. | Olivia | Shane | Female | [os@gmail.com](mailto:os@gmail.com) | Null | Null | Tin | P03 |
| 4. | Hannah | Ethan | Female | [he@gmail.com](mailto:he@gmail.com) | Null | Null | Joh | P04 |
| 5. | Megan | Cameron | Male | mec@gmail.com | Null | Null | JUl | P05 |
| 6. | Chloe | Tim | Male | [ct@gmail.com](mailto:ct@gmail.com) | Null | Null | CT | CT06 |
| 7. | Jess | David | Male | [jd@gmail.com](mailto:jd@gmail.com) | Null | Null | JD | JD07 |
| 8. | Caitlin | Andre | Female | [ca@gmail.com](mailto:ca@gmail.com) | Null | Null | CA | CA08 |
| 9. | Tallulah | Lawrence | Male | [tl@gmail.com](mailto:tl@gmail.com) | Null | Null | TL | TL09 |
| 10. | Ammaarah | Joshua | Female | [aj@gmail.com](mailto:aj@gmail.com) | Null | Null | AJ | AJ10 |
| 11. | Michelle | Johann | Female | [mj@gmail.com](mailto:mj@gmail.com) | Null | Null | MJ | MJ11 |
| 12. | Jenna | Ryan | Female | [jr@gmail.com](mailto:jr@gmail.com) | Null | Null | JR | JR12 |
| 13. | Haajarah | James | Male | [hj@gmail.com](mailto:hj@gmail.com) | Null | Null | HJ | HJ13 |
| 14. | Emma | Matt | Male | em@gmail.com | Null | Null | EM | EM14 |
| 15. | Tanja | Xavier | Male | [tx@gmail.com](mailto:tx@gmail.com) | Null | Null | TX | TX15 |
| 16. | Zoe | Calvin | Male | [zc@gmail.com](mailto:zc@gmail.com) | Null | Null | ZC | ZC16 |
| 17. | Maria | Isaac | Female | [mi@gmail.com](mailto:mi@gmail.com) | Null | Null | MI | MI17 |
| 18. | Leah | Armand | Female | [la@gmail.com](mailto:la@gmail.com) | Null | Null | LA | LA18 |
| 19. | Samantha | Luke | Female | [sl@gmail.com](mailto:sl@gmail.com) | Null | Null | SL | SL19 |
| 20. | Ella | Brendan | Female | [eb@gmail.com](mailto:eb@gmail.com) | Null | Null | EB | EB20 |
| 21. | Rachel | Corey | Female | [rc@gmail.com](mailto:rc@gmail.com) | Null | Null | RC | RC21 |
| 22. | Bianca | Thomas | Female | [bt@gmail.com](mailto:bt@gmail.com) | Null | Null | BT | BT22 |
| 23. | Kayla | Bosman | Female | [kb@gmail.com](mailto:kb@gmail.com) | Null | Null | KB | KB23 |
| 24. | Mavita | Bryan | Male | mb@gmail.com | Null | Null | MB | MB24 |
| 25. | Nina | Paci | Male | [np@gmail.com](mailto:np@gmail.com) | Null | Null | NP | NP25 |
| 26. | Danya | Fedinkgoeng | Female | [df@gmail.com](mailto:df@gmail.com) | Null | Null | DF | DF26 |
| 27. | Mailaika | Nkwabi | Male | mn@gmail.com | Null | Null | MN | MN27 |
| 28. | Lily | Errol | Female | [le@gmail.com](mailto:le@gmail.com) | Null | Null | LE | LE28 |
| 29. | Courtney | Treasure | Male | [ct@gmail.com](mailto:ct@gmail.com) | Null | Null | CT | CT29 |
| 30. | Kate | Emershan | Female | [ke@gmail.com](mailto:ke@gmail.com) | Null | Null | KE | KE30 |
| 31. | Sameera | Cowen | Male | [sc@gmail.com](mailto:sc@gmail.com) | Null | Null | SC | SC31 |
| 32. | Laila | Rowan | Female | [lr@gmail.com](mailto:lr@gmail.com) | Null | Null | LR | LR32 |
| 33. | Kendal | Quincy | Female | [kq@gmail.com](mailto:kq@gmail.com) | Null | Null | KQ | KQ33 |
| 34. | Amber | Kreeasen | Male | [ak@gmail.com](mailto:ak@gmail.com) | Null | Null | AK | AK34 |
| 35. | Mia | Sharief | Female | ms@gmail.com | Null | Null | MS | MS35 |
| 36. | Mishka | Sbusiso | Male | ms@gmail.com | Null | Null | MIS | MIS36 |

**3.4 Choice and Justification of Programming Language**

Python is a high-level and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Furthermore, python is also compatible with other programming languages such as Java and C++.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

**3.5 Advantages of the Proposed System**

The following advantages of the Proposed System are:

1. The ability of the proposed system to be less intrusive to potential malicious employee farmers who intends to hack the model in order to boycott monitoring
2. The ability to track, detect and prevent animal abuse by potential malicious employee farmers
3. The ability to monitor agricultural activities especially the environment, irrigation processes and farm cattle

**CHAPTER FOUR**

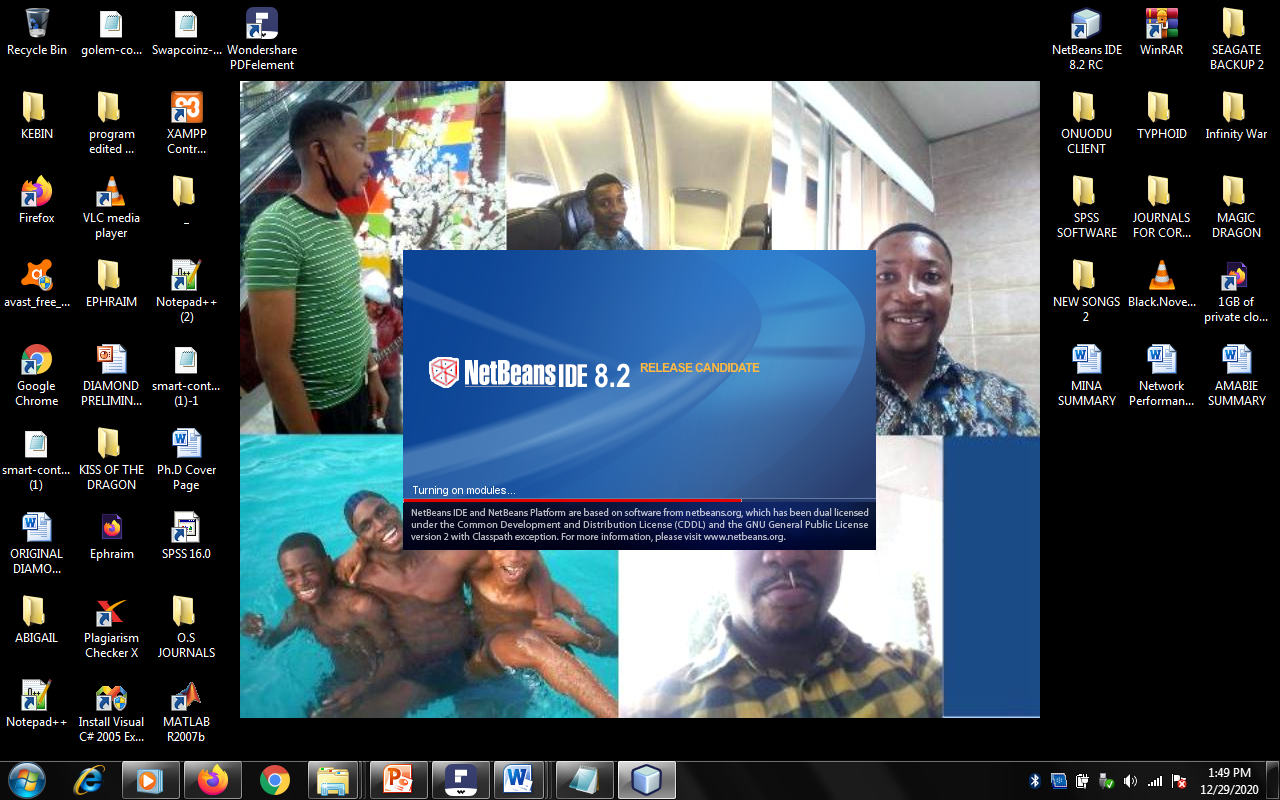
**RESULTS AND DISCUSSION**

**4.1 System Requirements**

The new system is a computerized model that emulates the decision-making ability of a farm manager. Requirements for the new system encompasses both hardware and software tools. The hardware tools include a PC (Personal Computer), Hard Disk Drive and Windows Operating System. The software tools include Windows Operating System, Internet Service Provider, Xampp Server, Python Modules and Interpreters and MySQL.

**4.2 Outputs and Discussion**

Figure 4.1 shows the initialization of our Python Project in Netbeans 8.2RC IDE. NetBeans is an integrated development environment (IDE) for Java and other Object-Oriented programming languages. NetBeans allows applications to be developed from a set of modular software components called modules. Applications based on NetBeans, including the NetBeans IDE, can be extended by third party developers. Figure 4.2 shows the welcome page of the new system. The welcome page consists of 3 navigation links namely the home, register and farm monitor. The register link navigates the user to a registration page (Figure 4.3), while the farm monitor link navigates the user to the test-input page in order to commence farm monitoring activities. Figure 4.4 shows the login page for registered users. The login page consists of fields for username and corresponding password. Figure 4.5 consists of the test-set input analysis which enables the user to key in essential information such as name, location, real-time monitoring request and so on. Figure 4.6 shows the real-time monitoring of the livestock and farmers.



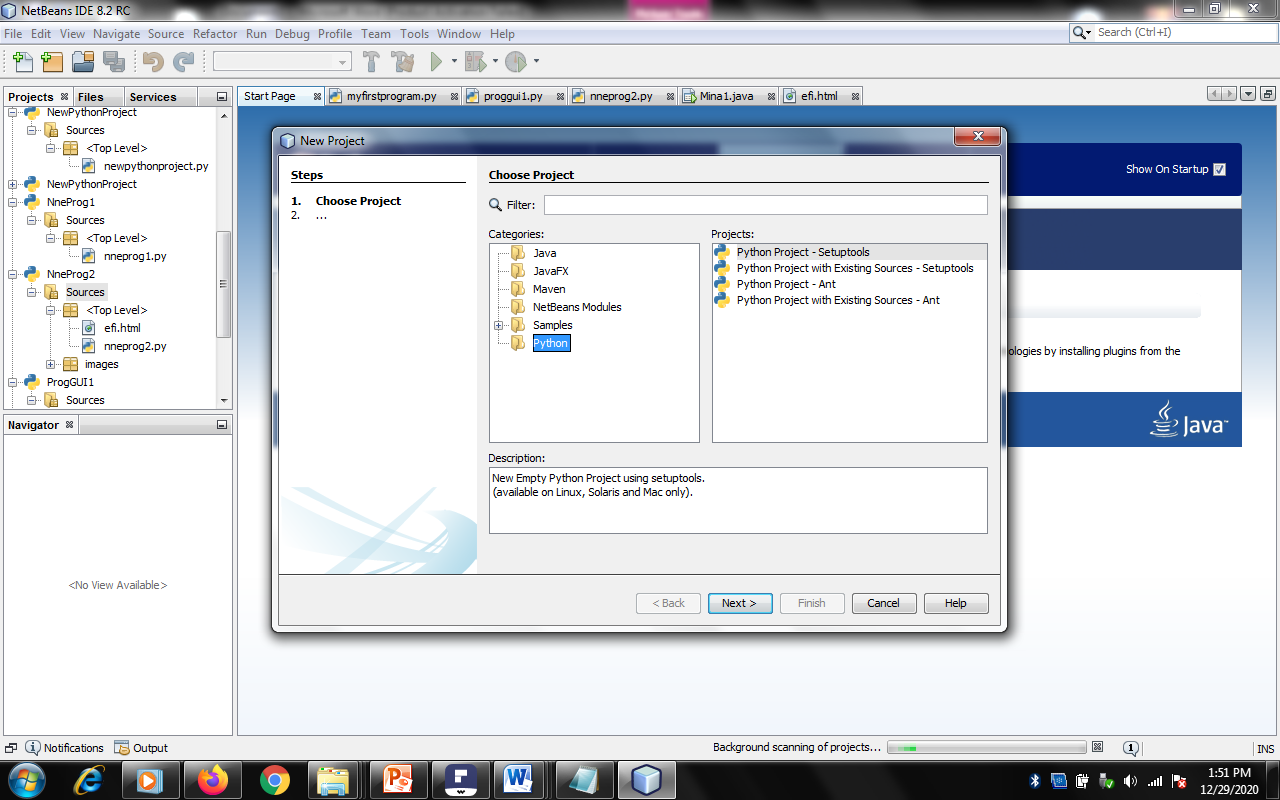


Figure 4.1: Initialization of Python Project in Netbeans 8.2RC IDE



Figure 4.2: Welcome Page

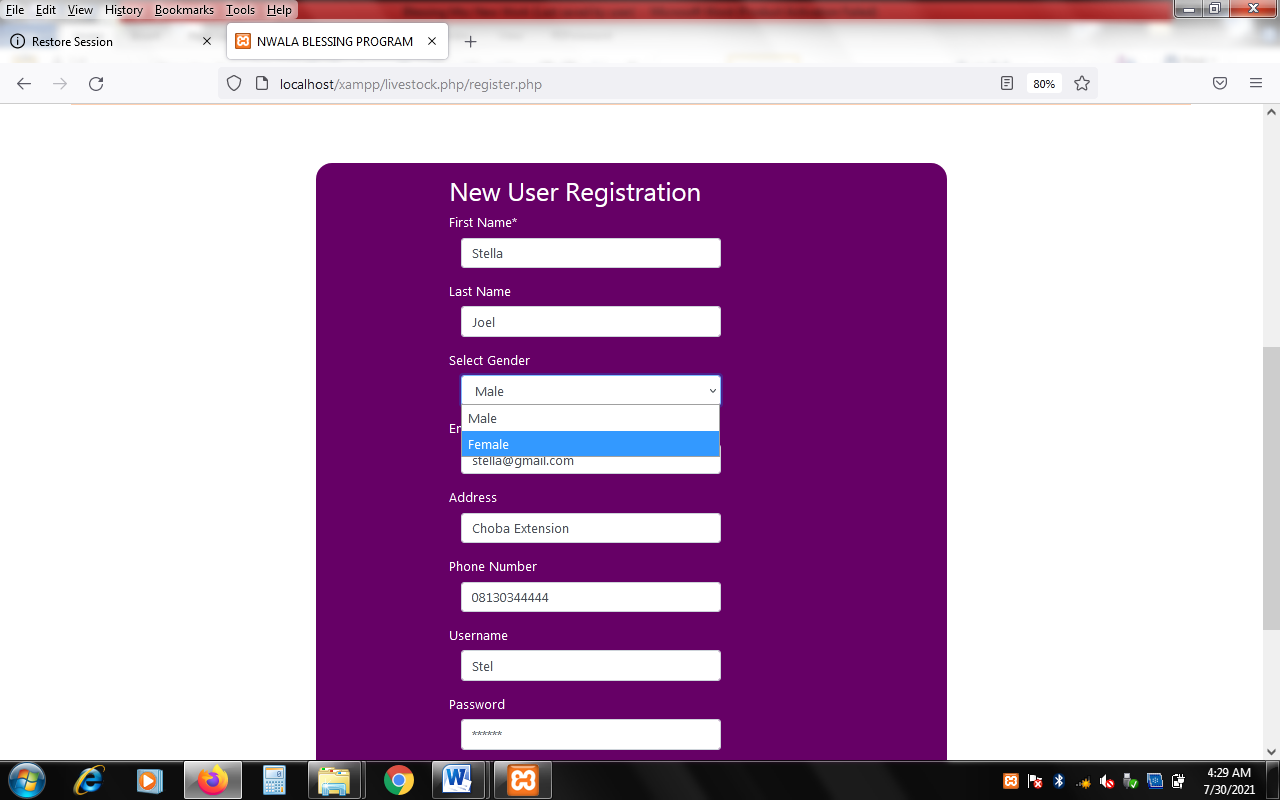


Figure 4.3: Registration Page for new User

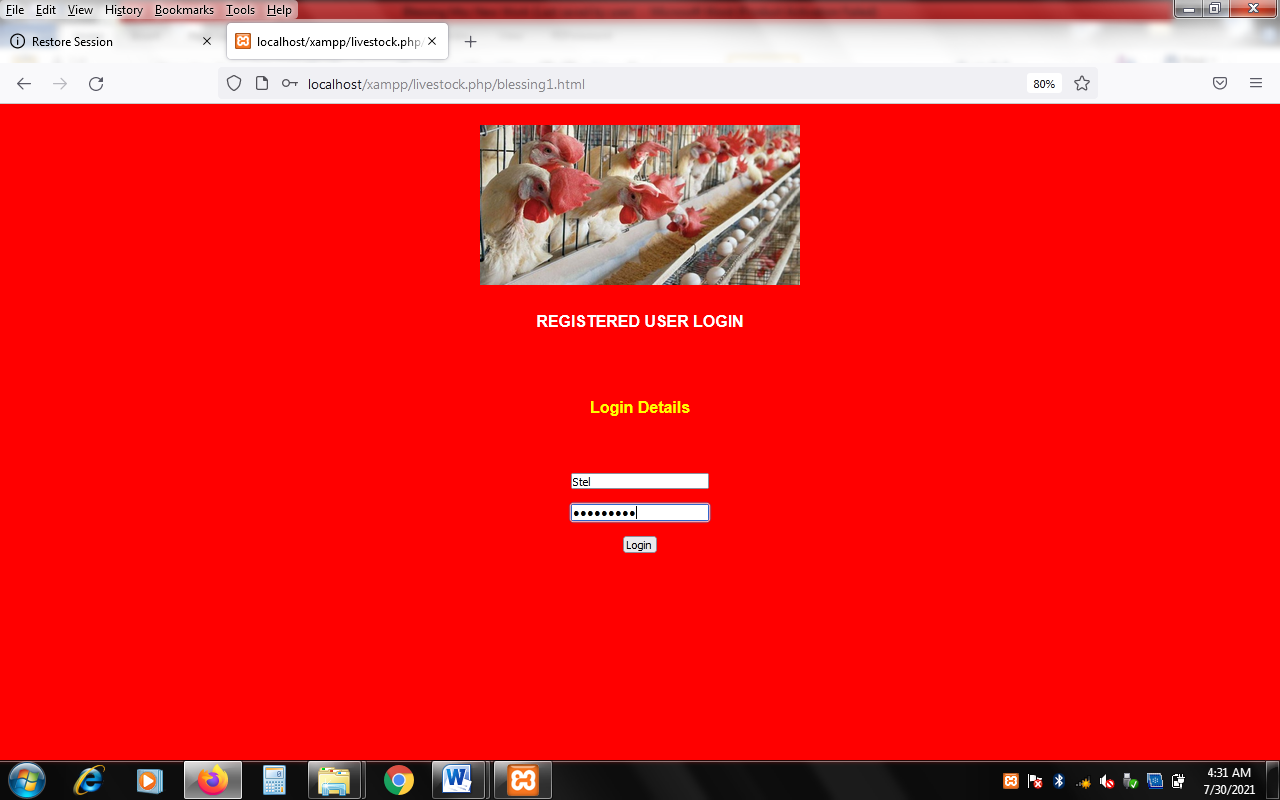


Figure 4.4: Login Page for Registered Users

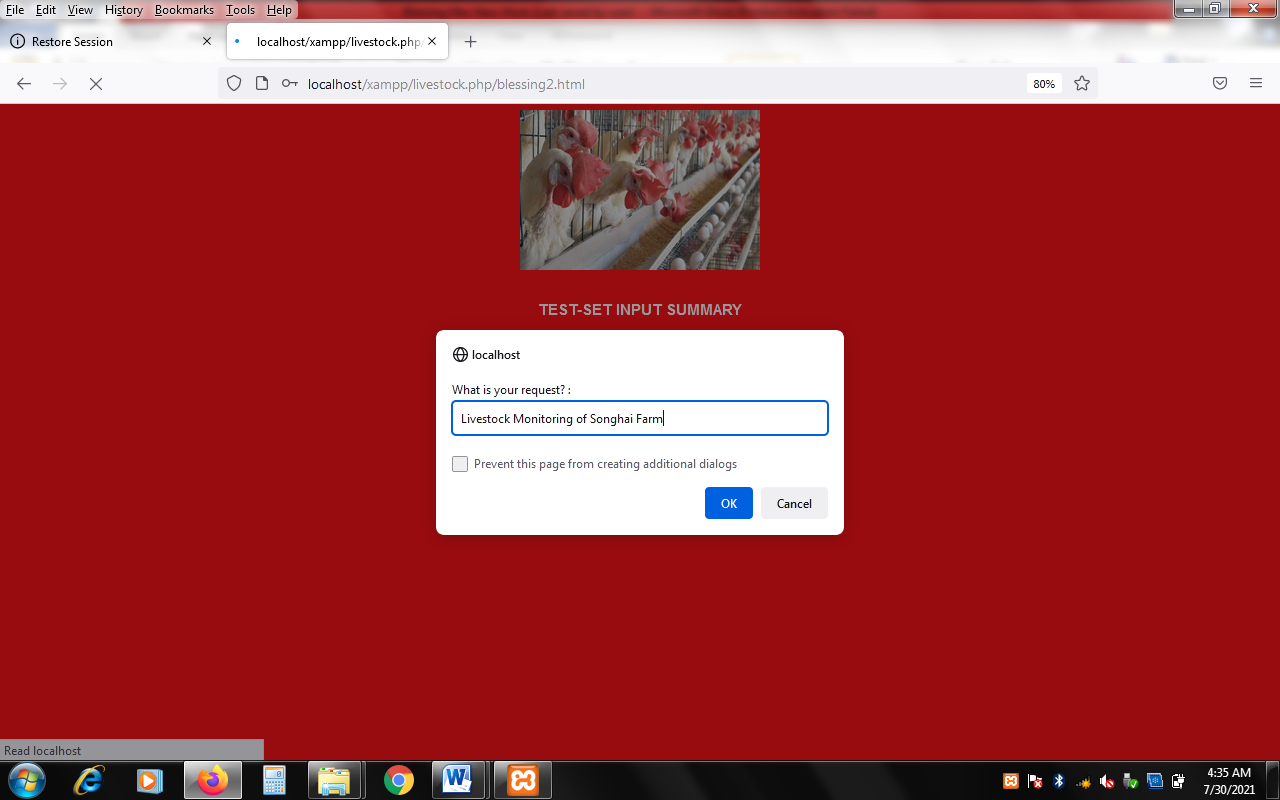


Figure 4.5: Test-Set Input Page

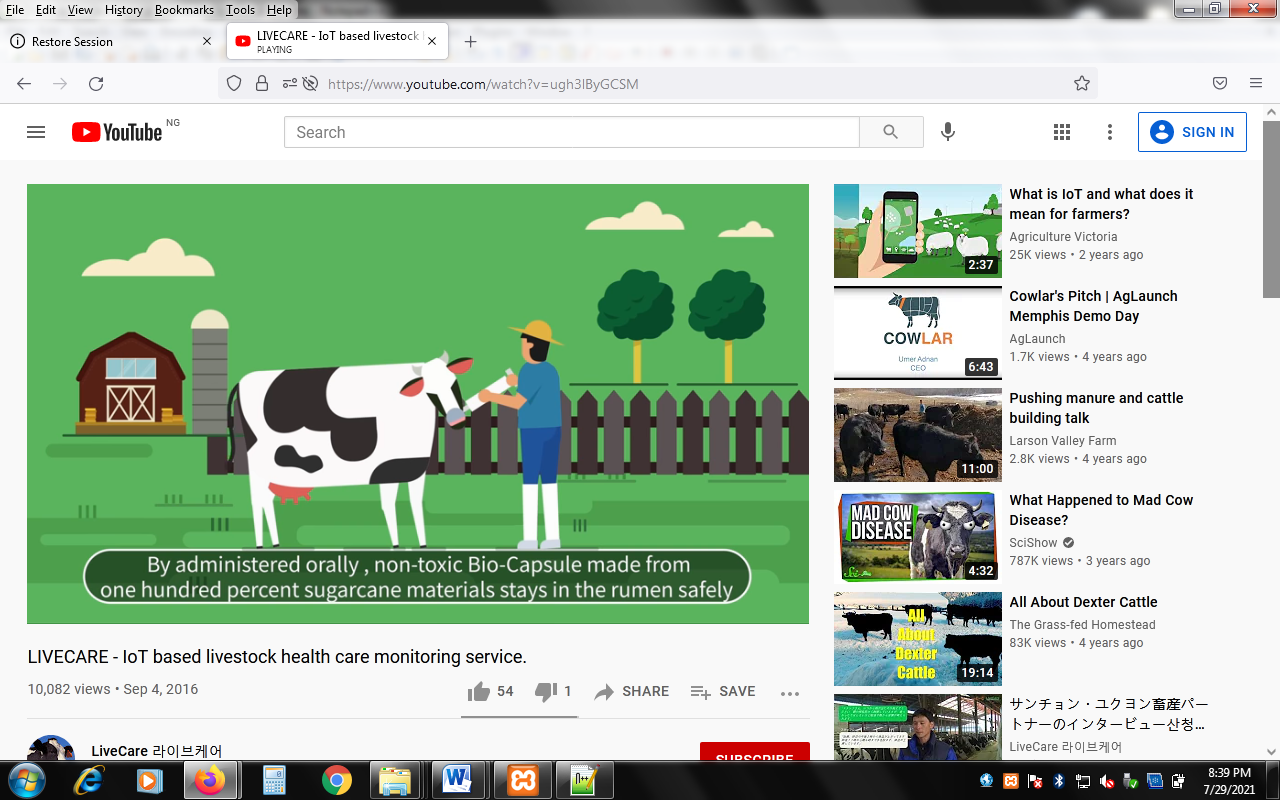


Figure 4.6: livestock monitoring and visualization

**4.2.1 Performance Evaluation**

Table 4.4: Detected Monitoring Possibility Calculation for the Proposed System

|  |  |
| --- | --- |
| TP  Actual value | FP |
| FN  Predicted value | TN |

Where TP = True Positive = 72

FP = False Positive = 10

FN = False Negative = 13

TN = True Negative = 5

Table 4.5: Confusion Matrix Table

|  |  |  |
| --- | --- | --- |
|  | No. of Possibilities for Successful Livestock Monitoring | No. of Possibilities for Unsuccessful Livestock Monitoring |
| Occurring | 72 | 5  77 |
| Not Occurring | 10 | 13  13 |
| Total Sample | 82 | 18 |

To obtain overall accuracy of the Proposed System

Accuracy = No. of correct possibilities

(4.3) (4.1)

Total number of possibilities made

= TP + TN

TP + TN + FP + FN

= 72 + 17/72 + 5 + 10 + 13

= 89/100 = 0.77 = 77%

Misclassification rate (a.k.a. classification error) = (FP+FN) / (TP+TN+FP+FN) (4.4)

Or 1 – Accuracy of the System = 1 -89% = 0.11 = 11%

Table 4.6: Performance Evaluation Matrix table for the Proposed System

|  |  |  |
| --- | --- | --- |
|  | Accuracy | Classification  Error |
| Successful Monitoring | 89% | 11% |
| Unsuccessful Monitoring |

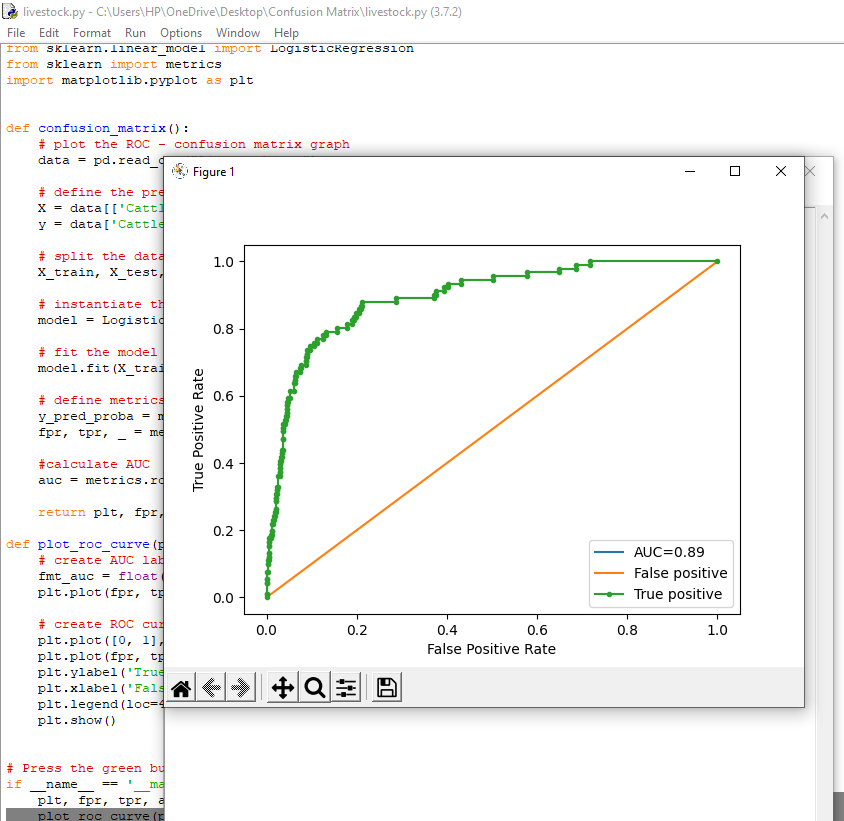


Figure 4.7: ROC curve of the proposed system

**4.4 Discussion of Results**

Figure 4.1 showed the Netbeans IDE initialization page for Python Modules. Python is an Object-Oriented Programming Language. Figure 4.2 showed the Welcome Page of the New System. Figures 4.3 and 4.4 showed the new user registration and login pages respectively. Figure 4.5 showed the Test-Set Input page, while Figure 4.6 showed the real-time monitoring of livestock and farmers. Furthermore, the study adopted Confusion Matrix for performance evaluation of the existing and new systems. The mentioned method is required in order to know the strength of the new system especially in terms of usage and result accuracy. In order to know the strength of the new system; a confusion matrix was obtained, for the accuracy and classification error to be computed. The confusion matrix gives a matrix output, and further describes the performance of the model.

The existing system used calibrated and uncalibrated measurement method and obtained an accuracy rate of 59%. The following parameters made up the actual and predicted values of the confusion matrix for the proposed system.

1. “true positive (TP)” for correctly predicted event values.
2. “false positive (FP)” for incorrectly predicted event values.
3. “true negative (TN)” for correctly predicted no-event values.
4. “false negative (FN)” for incorrectly predicted no-event values.

Where

TP = 72

FP = 10

FN = 13

TN = 5

No. of Possibilities for Successful Livestock Monitoring occurring = 72

No. of Possibilities for Successful Livestock Monitoring not occurring = 72

The proposed system obtained an accuracy rate of 89% with a minimal classification error rate of 11%. This was further illustrated with the ROC curve of the proposed system which was derived from our datasets and further implemented with python programming (Figure 4.7).

**CHAPTER FIVE**

**SUMMARY, CONCLUSION AND RECOMMENDATIONS**

**5.1 Summary**

The study opined that livestock farming is a component of modern agriculture that has been practiced in many cultures since humanity's transition to farming from hunter-gatherer lifestyles. Animal husbandry practices have varied widely across cultures and time periods, and continue to play a major economic and cultural role in numerous communities. Livestock farming practices have largely shifted to intensive animal farming, sometimes referred to as "factory farming"; over 99% of livestock in most countries are now raised in this way. Intensive animal farming increases the yield of the various commercial outputs, but has also led to negative impacts on animal welfare, the environment, and public health. In particular, livestock especially beef, dairy and sheep stocks, have out-sized influence on greenhouse gas emissions from agriculture. Animal-rearing originated during the cultural transition to settled farming communities from hunter-gatherer lifestyles. Animals are domesticated when their breeding and living conditions are controlled by humans. Over time, the collective behavior, lifecycle and physiology of livestock have changed radically.

**5.2 Conclusion**

In this study, we developed an Improved Model for Smart Livestock Monitoring and Management. Fuzzy Logic and Recurrent Neural Network were technologies that powered the performance accuracy of the developed system. A recurrent neural network is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior.

**5.3 Recommendations**

The study also suggested the importance of automated livestock monitoring to growing agriculture in Nigeria. This because; the goal of smart agriculture encompasses a decision making support system for optimized monitoring. Smart farming deems it necessary to address the issues of population growth, climate change and labor that has gained a lot of technological attention, from planting and watering of crops to health and harvesting.

**5.4 Contributions to Knowledge**

The study contributed the following to enhancing smart agriculture and livestock monitoring:

1. An improved machine-learning model for automated livestock monitoring and management
2. A unique real-time visualization monitoring system for optimizing trust and security in livestock monitoring and management.
3. An enhanced framework for accurately tracking the health status of a suspected animal

**5.5 Limitations of the Study**

The developed model in the study was only simulated and not tested with real hardware IoT sensors. This was as a result of scarce resources during the study period

**5.6 Future Work**

The study intends to expand its scope to monitoring and security of Food Supply Chain Management processes in Nigeria in order to optimize food security and traceability.

**References**

Abraham B.E. (2019), Machine Learning Approach for Agricultural IOT, *International Journal*

*of Recent Technology & Engineering (IJRTE), 7(6), 383 – 392*

Ashalatha K., Simin C., Raluca M. & Cristina S. (2012), Architecture Modeling & Formal

Analysis of Intelligent Multi-Agent Systems, *An Article published to Malardalen University, Vasteras Sweden*

Athanasios V., Nikolaos D., Anastasios D. & Eftychios P. (2017), Deep Learning for Computer

Vision: A Brief Review, *Hindawi Computational Intelligence & Neuro-Science, Volume 2018, Article ID 7068349, https://doi.org/10./155/2018/7068349*

Ayman S., Osamah K., & Ghaida A. (2016), An Adaptive Intelligent Alarm System for

Wireless Sensor Network, *Indonesian Journal of Electrical Engineering & Computer Science, 15(1), 142 – 147*

Ayon D. (2016), Machine Learning Algorithms: A Review, *International Journal of Computer*

*Science & Information Technologies, 7(3), 1174 – 1179*

Baki C., Kenneth H., Maria G., Paul S., Anette H. (2018), CASE: A framework for computer

supported outbreak detection, *International Journal of Computer Applications (IJCA), 8(7), 44 – 53*

Basma E., Sherine E., & Hussein E. (2014), Independent Living for persons with disabilities

& elderly people using smart home technology, *International Journal of Application or Innovation in Engineering & Management (IJAIEM), 3(4), 11 – 28*

Bethany K., Alex L., & Jonathan C. (2017), Evolution of Smart Homes for the Elderly, *ACM*

*978-1-4503-4914-7/17/04, doi:http://dx.doi.org/10.1145/3041021.3054928*

Bradley, A. W., & Murphy, G. C. (2011), Supporting Software History Exploration,

*Proceedings of the 8th Working Conference on Mining Software Repositories, Waikiki, Honolulu, HI, USA, ACM Press, 193-202*

Bagbaikpe K.H. (2019), A Review of Machine Learning & IOT in Smart Transportation,

*Future Internet, MDPI Journal*

Banjol S.D. (2019), Clot-Net: A Scalable Cognitive IoT based Smart City Network &

Architecture, *Human Centric Computing & Information Sciences, https://doi.org/10.1186/5/3673-019-0190-9*

Chatley, R. & Jones, L. (2018), Diggit: Automated Code Review via Software Repository

Mining, *Proceeding of the IEEE 25th International Conference on Software Analysis, Evolution & Reengineering (SANER), IEEE Computer Society, 567-571*

Chen, H. (2016). Applications of Fuzzy Logic in Data Mining Process. In Z. H. Bai Y., *Advanced Fuzzy Logic Technologies in Industrial Applications. Advances in Industrial Control.* Springer: London.

Chibroma A. (2017), Structured Generative Models using the IoT, *Proceedings of the*

*31st International Conference on Machine Learning, Beijing, China, JMLR: W & CP, V.32, arXiv:1401.0514v2[cs.PL], 1 – 14*

Dagmawi N., Paolo S., Nicola F. & Aldo D. (2019), Reasoning in Multi-Agent based Smart

Homes: A Systematic Literature Review, *https://www.researchgate.net/publication/330836390*

Daniel S., Fabian L., & Ingo T. (2018), Agent-based M & S of Individual Elderly Care

Decision-Making, *Proceedings of the 2018 Winter Simulation Conference*

Danilo R., Shakir M., Ivo D., Karol G., & Daan W. (2016), One-shot Generalization in Deep

Generative Models, *Proceedings of the International Conference on Machine Learning, New York, NY, USA, 2016, JMLR, W&CP, VOL. 48, 1 – 9*

David A. & Graham P. (2012), Design Science in Decision Support Systems Research: An

Assessment using the Hevner, March Park, & Ram Guidelines, *Journal of the Association for Information Systems (JAIS), 13(11), 923 – 949*

Debajyoti P., Suree F., Vajirasak V., & Borworn P. (2018), Analyzing the Elderly Users’

Adoption of Smart-Home Services, *IEEE Access, 2018.2869599, doi:10.1109*

Dyer, R., Nguyen, H. A., Rajan, H. & Nguyen,T. N. (2015). Boa: Ultra- Large-Scale Software

Repository & Source Code Mining, *ACM Transactions on Software Engineering & Methodology (TOSEM), 25(1)*

Eissa A. (2019), Smart Sustainable Agriculture (SSA) Solution underpinned by Internet of

Things (IOT) & Artificial Intelligence (AI), *International Journal of Advanced Computer Science & Applications, 10(5), 93 – 102*

Fatima J.I. (2019), Machine Learning in IOT Security: Current Solutions & Future Challenges,

*Arxiv:1904.05735v1[cs.CR]*

Fontios Y.G. (2019), A Review of Machine Learning & IOT in Smart Transportation, *Future*

*Internet, MDPI Journal*

Ekong V. (2013), A Fuzzy Inference System for predicting Depression risk levels,

*African Journal of Mathematics & Computer Science Research, 6(10), 197 – 204*

Eric N., Akihiro M., Yee T., Dilan G., Bolaji L., (2019), Do Deep Generative Models know what

they don’t know? *Published as a Conference paper at ICLR 2019*

Lamik I., Anthony B., Mihaela S., & Charlotte D (2020), *Deep Generative Models for 3D*

*Linker Design, J.Chem. Inf. Model. 2020, 60, 1983 – 1995*

Frederick I., Jones O. & Cecilia F. (2017), A Modified Approach to Artificial Intelligence (AI)

for Activity Guidance, *International Journal of Computer Application (IJCA), 8(9), 1 – 5*

Fred H., Niu D., Haolan C., Kunfeng L., Yancheng H., & Yu. X. (2019), A Deep Generative

Approach to Search Extrapolation & Recommendation, *in the 25th ACM SIGKDD Conference on knowledge discovery & data mining (KDD ’19), August 4 – 8, 2019, https://doi.org/10.1/45/3292500.3330786*

Frey I. & Osborne A. (2013), The impact of Artificial Intelligence in the Modern Century,

*International Journal of Engineering Technology (IJET), 4(9), 3 – 9*

George M., Amritanshu A., & Tim M. (2018), Finding Trends in Software Research, *IEEE*

*Transactions in Software Engineering, arXiv:1608.08100v10[cs.SE] 3rd October, 2018, 1 – 12*

Gbenga M.M (2019), Machine Learning in IOT Security: Current Solutions & Future

Challenges, *Arxiv:1904.05735v1[cs.CR]*

Hannah M. & Julie S. (2019), A Review of Age Friendly Virtual Assistive Technologies &

their Effect on Daily Living for Carers & Dependent Adults, *HealthCare 2019, 7, 49; doi:10.3390/healthcare7010049*

Todor T. & Juri S. (2017), Expert System for Milk & Animal Monitoring, *International*

*Journal of Advanced Computer Applications, 7(4), 22 – 34*

**Appendix A**

**Source Codes**

public void actionPerformed(python.awt.event.ActionEvent evt) {

jButton1ActionPerformed(evt);

}

});

getContentPane().add(jButton1);

jButton1.setBounds(260, 440, 300, 50);

jButton2.setFont(new python.awt.Font("Tahoma", 1, 18)); // NOI18N

jButton2.setText("REGISTER");

jButton2.addActionListener(new python.awt.event.ActionListener() {

public void actionPerformed(python.awt.event.ActionEvent evt) {

jButton2ActionPerformed(evt);

}

});

getContentPane().add(jButton2);

jButton2.setBounds(40, 440, 130, 50);

pack();

}// </editor-fold>

break;

}

}

} catch (ClassNotFoundException ex) { python.util.logging.Logger.getLogger(PythonWelcome.class.getName()).log(python.util.logging.Level.SEVERE, null, ex);

} catch (InstantiationException ex) { python.util.logging.Logger.getLogger(PythonWelcome.class.getName()).log(python.util.logging.Level.SEVERE, null, ex);

} catch (IllegalAccessException ex) { python.util.logging.Logger.getLogger(PythonWelcome.class.getName()).log(python.util.logging.Level.SEVERE, null, ex);

} catch (python.swing.UnsupportedLookAndFeelException ex) { python.util.logging.Logger.getLogger(PythonWelcome.class.getName()).log(python.util.logging.Level.SEVERE, null, ex);

}

//</editor-fold

/\* Create and display the form \*/

python.awt.EventQueue.invokeLater(new Runnable() {

public void run() {

new PythonWelcome().setVisible(true);

}

});

}

// Variables declaration - do not modify

private python.swing.JButton jButton1;

private python.swing.JButton jButton2;

private python.swing.JLabel jLabel1;

private python.swing.JLabel jLabel2;

private python.swing.JLabel jLabel3;

private python.swing.JLabel jLabel4;

private python.swing.JLabel jLabel5;

# To change this license header, choose License Headers in Project Properties.

# To change this template file, choose Tools | Templates

# and open the template in the editor.

frame = JFrame("Go to PythonWelcome")

frame.setDefaultCloseOperation(JFrame.EXIT\_ON\_CLOSE)

frame.setLocation(100,100)

frame.setSize(300,200)

frame.setVisible(True)

<python

function sanitize($input){

$input = mysqli\_real\_escape\_string($GLOBALS['connect\_database'], $input);

return $input;

}

function mysqli\_result($res, $row=0, $col=0){

$numrows = mysqli\_num\_rows($res);

if($numrows && $row <=($numrows-1) && $row >=0){

mysqli\_data\_seek($res,$row);

$resrow = (is\_numeric($col)) ? mysqli\_fetch\_row($res) : mysqli\_fetch\_assoc($res);

if (isset($resrow[$col])){

return $resrow[$col];

}

}

return false;

}

function user\_exists($username){

$query = mysqli\_query($GLOBALS['connect\_database'], "SELECT COUNT(user\_id) FROM hilivis.registered\_users WHERE username= '$username' ");

return mysqli\_result($query, 0);

}

function email\_exists($email){

$query = mysqli\_query($GLOBALS['connect\_database'], "SELECT COUNT(user\_id) FROM

hilivis.registered\_users WHERE email = '$email' ");

return mysqli\_result($query, 0);

}

function register\_user($first\_name, $last\_name, $gender, $email, $address, $phone\_number, $username, $password){

//echo 'You\'re eligible to register!';

mysqli\_query($GLOBALS['connect\_database'], "

INSERT INTO hilivis.registered\_users(first\_name, last\_name,

gender, email, address, phone\_number, username, password, timestamp) VALUES('$first\_name', '$last\_name', '$gender', '$email', '$address', '$phone\_number',

'$username', '$password', UNIX\_TIMESTAMP()) ");

}

function user\_name\_matches\_phone($username, $phone\_number){

$username = sanitize($username);

$phone\_number = sanitize($phone\_number);

$query = mysqli\_query($GLOBALS['connect\_database'], "SELECT COUNT(user\_id) FROM hilivis.registered\_users WHERE username= '$username' AND phone\_number = '$phone\_number' ");

return mysqli\_result($query, 0);

}

function retrieve\_recovery\_data($username, $phone\_number){

$user\_data = array();

$new\_password = 'Akjffr235ssxz090';

mysqli\_query($GLOBALS['connect\_database'], "UPDATE hilivis.registered\_users SET password = '$new\_password' ");

$query = mysqli\_query($GLOBALS['connect\_database'], "SELECT email, password FROM hilivis.registered\_users WHERE username= '$username' AND phone\_number = '$phone\_number' ");

while($row = mysqli\_fetch\_assoc($query)){

$user\_data[] = array(

'password' => $row['password'],

'email' => $row['email']

);

}

return $user\_data;

}

?>

<!DOCTYPE html>

<html>

<head>

<html>

<head>

<meta charset="utf-8">

<title>NWALA BLESSING PROGRAM</title>

<link href="css/bootstrap.min.css" rel="stylesheet"/>

<!--Costomer css-->

<link rel="stylesheet" href="css/index.css">

<link rel="stylesheet" type="text/css" href="css/animation/animate.min.css">

</head>

<body>

<nav class="navbar navbar-expand-sm bg-dark navbar-dark"

style="height: 70px;">

<div class="container-fluid">

<a class="navbar-brand animate fadeInDown" href="index.php" style="animation-duration: 2s;">

<!-- <img src="images/logo\_school.png" alt="Cinque Terre" width="70"> -->

<span class="name">IMPROVED MONITORING MODEL FOR SMART LIVESTOCK FARMING<br>

</span>

</a>

<ul class="navbar-nav navbdfar-right">

<li class="nav-item">

<a class="nav-link" id="nav-link" href="index.php">Home</a>

</li>

<li class="nav-item">

<a class="nav-link" id="nav-link" href="register.php">Register </a>

</li>

<li class="nav-item">

<a class="nav-link" id="nav-link" href="blessing1.html">Farm Monitor</a></li>

</ul>

</div>

</nav>

</div> <br><br>

<!-- End header section -->

<?php

include "second\_slider.php";

?>

<!-- <center>

<img src="images/lady\_computer.jpg" alt="pattern pix" width="1000" height="500"

style="border-radius: 20px;" />

</center -->

<BR>

<center>

<font color="black" size="5" face="Arial">

<!-- #region Jssor Slider Begin -->

<!-- Generator: Jssor Slider Maker -->

<!-- Source: https://www.jssor.com -->

<div style="">

<script src="slider/js/jssor.slider.min.js" type="text/javascript"></script>

<script type="text/javascript">

jssor\_1\_slider\_init = function() {

var jssor\_1\_SlideshowTransitions = [

{$Duration:500,$Delay:30,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:2049,$Easing:$Jease$.$OutQuad},

{$Duration:500,$Delay:80,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Easing:$Jease$.$OutQuad},

{$Duration:1000,x:-0.2,$Delay:40,$Cols:12,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Assembly:260,$Easing:{$Left:$Jease$.$InOutExpo,$Opacity:$Jease$.$InOutQuad},$Opacity:2,$Outside:true,$Round:{$Top:0.5}},

{$Duration:2000,y:-1,$Delay:60,$Cols:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Easing:$Jease$.$OutJump,$Round:{$Top:1.5}},

{$Duration:1200,x:0.2,y:-0.1,$Delay:20,$Cols:8,$Rows:4,$Clip:15,$During:{$Left:[0.3,0.7],$Top:[0.3,0.7]},$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:260,$Easing:{$Left:$Jease$.$InWave,$Top:$Jease$.$InWave,$Clip:$Jease$.$OutQuad},$Round:{$Left:1.3,$Top:2.5}}

];

var jssor\_1\_options = {

$AutoPlay: 1,

$SlideshowOptions: {

$Class: $JssorSlideshowRunner$,

$Transitions: jssor\_1\_SlideshowTransitions,

$TransitionsOrder: 1

},

$ArrowNavigatorOptions: {

$Class: $JssorArrowNavigator$

},

$BulletNavigatorOptions: {

$Class: $JssorBulletNavigator$

}

};

var jssor\_1\_slider = new $JssorSlider$("jssor\_1", jssor\_1\_options);

/\*#region responsive code begin\*/

var MAX\_WIDTH = 1400;

function ScaleSlider() {

var containerElement = jssor\_1\_slider.$Elmt.parentNode;

var containerWidth = containerElement.clientWidth;

if (containerWidth) {

var expectedWidth = Math.min(MAX\_WIDTH || containerWidth, containerWidth);

jssor\_1\_slider.$ScaleWidth(expectedWidth);

}

else {

window.setTimeout(ScaleSlider, 30);

}

}

ScaleSlider();

$Jssor$.$AddEvent(window, "load", ScaleSlider);

$Jssor$.$AddEvent(window, "resize", ScaleSlider);

$Jssor$.$AddEvent(window, "orientationchange", ScaleSlider);

/\*#endregion responsive code end\*/

};

</script>

<style>

/\* jssor slider loading skin spin css \*/

.jssorl-009-spin img {

animation-name: jssorl-009-spin;

animation-duration: 1.6s;

animation-iteration-count: infinite;

animation-timing-function: linear;

}

@keyframes jssorl-009-spin {

from {

transform: rotate(0deg);

}

to {

transform: rotate(360deg);

}

}

.jssorb053 .i {position:absolute;cursor:pointer;}

.jssorb053 .i .b {fill:#fff;fill-opacity:0.5;}

.jssorb053 .i:hover .b {fill-opacity:.7;}

.jssorb053 .iav .b {fill-opacity: 1;}

.jssorb053 .i.idn {opacity:.3;}

.jssora093 {display:block;position:absolute;cursor:pointer;}

.jssora093 .c {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093 .a {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093:hover {opacity:.8;}

.jssora093.jssora093dn {opacity:.6;}

.jssora093.jssora093ds {opacity:.3;pointer-events:none;}

</style>

<div id="jssor\_1" style="position:relative;margin:0 auto;top:0px;left:0px;width:980px;height:380px;overflow:hidden;visibility:hidden;">

<!-- Loading Screen -->

<div data-u="loading" class="jssorl-009-spin" style="position:absolute;top:0px;left:0px;width:100%;height:100%;text-align:center;background-color:rgba(0,0,0,0.7);">

<img style="margin-top:-19px;position:relative;top:50%;width:38px;height:38px;" src="../svg/loading/static-svg/spin.svg" />

</div>

<div data-u="slides" style="cursor:default;position:relative;top:0px;left:0px;width:980px;height:380px;overflow:hidden;">

<div>

<img data-u="image" src="images/pix1.jpg" />

</div>

<div>

<img data-u="image" src="images/pix2.jpg" />

</div>

<div>

<img data-u="image" src="images/pix3.jpg" />

</div>

<div style="background-color:#ff7c28;">

<div style="position:absolute;top:50px;left:50px;width:450px;height:62px;z-index:0;font-size:16px;color:#000000;line-height:24px;text-align:left;padding:5px;box-sizing:border-box;">Photos in this slider are to demostrate jssor slider,<br />

which are not licensed for any other purpose.

</div>

</div>

</div>

<!-- Bullet Navigator -->

<div data-u="navigator" class="jssorb053" style="position:absolute;bottom:12px;right:12px;" data-autocenter="1" data-scale="0.5" data-scale-bottom="0.75">

<div data-u="prototype" class="i" style="width:16px;height:16px;">

<svg viewBox="0 0 16000 16000" style="position:absolute;top:0;left:0;width:100%;height:100%;">

<path class="b" d="M11400,13800H4600c-1320,0-2400-1080-2400-2400V4600c0-1320,1080-2400,2400-2400h6800 c1320,0,2400,1080,2400,2400v6800C13800,12720,12720,13800,11400,13800z"></path>

</svg>

</div>

</div>

<!-- Arrow Navigator -->

<div data-u="arrowleft" class="jssora093" style="width:50px;height:50px;top:0px;left:30px;" data-autocenter="2" data-scale="0.75" data-scale-left="0.75">

<svg viewBox="0 0 16000 16000" style="position:absolute;top:0;left:0;width:100%;height:100%;">

<circle class="c" cx="8000" cy="8000" r="5920"></circle>

<polyline class="a" points="7777.8,6080 5857.8,8000 7777.8,9920 "></polyline>

<line class="a" x1="10142.2" y1="8000" x2="5857.8" y2="8000"></line>

</svg>

</div>

<div data-u="arrowright" class="jssora093" style="width:50px;height:50px;top:0px;right:30px;" data-autocenter="2" data-scale="0.75" data-scale-right="0.75">

<svg viewBox="0 0 16000 16000" style="position:absolute;top:0;left:0;width:100%;height:100%;">

<circle class="c" cx="8000" cy="8000" r="5920"></circle>

<polyline class="a" points="8222.2,6080 10142.2,8000 8222.2,9920 "></polyline>

<line class="a" x1="5857.8" y1="8000" x2="10142.2" y2="8000"></line>

</svg>

</div>

</div>

<script type="t JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:2049,$Easing:$Jease$.$OutQuad},

{$Duration:500,$Delay:80,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Easing:$Jease$.$OutQuad},

{$Duration:1000,x:-0.2,$Delay:40,$Cols:12,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Assembly:260,$Easing:{$Left:$Jease$.$InOutExpo,$Opacity:$Jease$.$InOutQuad},$Opacity:2,$Outside:true,$Round:{$Top:0.5}},

{$Duration:2000,y:-1,$Delay:60,$Cols:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Easing:$Jease$.$OutJump,$Round:{$Top:1.5}},

{$Duration:1200,x:0.2,y:-0.1,$Delay:20,$Cols:8,$Rows:4,$Clip:15,$During:{$Left:[0.3,0.7],$Top:[0.3,0.7]},$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:260,$Easing:{$Left:$Jease$.$InWave,$Top:$Jease$.$InWave,$Clip:$Jease$.$OutQuad},$Round:{$Left:1.3,$Top:2.5}}

];

var jssor\_1\_options = {

$AutoPlay: 1,

$SlideshowOptions: {

$Class: $JssorSlideshowRunner$,

$Transitions: jssor\_1\_SlideshowTransitions,

$TransitionsOrder: 1

},

$ArrowNavigatorOptions: {

$Class: $JssorArrowNavigator$

},

$BulletNavigatorOptions: {

$Class: $JssorBulletNavigator$

}

};

var jssor\_1\_slider = new $JssorSlider$("jssor\_1", jssor\_1\_options);

/\*#region responsive code begin\*/

var MAX\_WIDTH = 1400;

function ScaleSlider() {

var containerElement = jssor\_1\_slider.$Elmt.parentNode;

var containerWidth = containerElement.clientWidth;

if (containerWidth) {

var expectedWidth = Math.min(MAX\_WIDTH || containerWidth, containerWidth);

jssor\_1\_slider.$ScaleWidth(expectedWidth);

}

else {

window.setTimeout(ScaleSlider, 30);

}

}

ScaleSlider();

$Jssor$.$AddEvent(window, "load", ScaleSlider);

$Jssor$.$AddEvent(window, "resize", ScaleSlider);

$Jssor$.$AddEvent(window, "orientationchange", ScaleSlider);

/\*#endregion responsive code end\*/

};

</script>

<style>

/\* jssor slider loading skin spin css \*/

.jssorl-009-spin img {

animation-name: jssorl-009-spin;

animation-duration: 1.6s;

animation-iteration-count: infinite;

animation-timing-function: linear;

}

@keyframes jssorl-009-spin {

from {

transform: rotate(0deg);

}

to {

transform: rotate(360deg);

}

}

.jssorb053 .i {position:absolute;cursor:pointer;}

.jssorb053 .i .b {fill:#fff;fill-opacity:0.5;}

.jssorb053 .i:hover .b {fill-opacity:.7;}

.jssorb053 .iav .b {fill-opacity: 1;}

.jssorb053 .i.idn {opacity:.3;}

.jssora093 {display:block;position:absolute;cursor:pointer;}

.jssora093 .c {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093 .a {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093:hover {opacity:.8;}

.jssora093.jssora093dn {opacity:.6;}

.jssora093.jssora093ds {opacity:.3;pointer-events:none;}

</style>

<div id="jssor\_1" style="position:relative;margin:0 auto;top:0px;left:0px;width:980px;height:380px;overflow:hidden;visibility:hidden;">

<!-- Loading Screen -->

<div data-u="loading" class="jssorl-009-spin" style="position:absolute;top:0px;left:0px;width:100%;height:100%;text-align:center;background-color:rgba(0,0,0,0.7);">

<img style="margin-top:-19px;position:relative;top:50%;width:38px;height:38px;" src="../svg/loading/static-svg/spin.svg" />

</div>

<div data-u="slides" style="cursor:default;position:relative;top:0px;left:0px;width:980px;height:380px;overflow:hidden;">

<div>

<img data-u="image" src="images/pix1.jpg" />

</div>

<div>

<img data-u="image" src="images/pix2.jpg" />

</div>

<div>

<img data-u="image" src="images/pix3.jpg" />

</div>

<div style="background-color:#ff7c28;">

<div style="position:absolute;top:50px;left:50px;width:450px;height:62px;z-index:0;font-size:16px;color:#000000;line-height:24px;text-align:left;padding:5px;box-sizing:border-box;">Photos in this slider are to demostrate jssor slider,<br />

which are not licensed for any other purpose.

</div>

</div>

</div>

<!-- Bullet Navigator -->

<div data-u="navigator" class="jssorb053" style="position:absolute;bottom:12px;right:12px;" data-autocenter="1" data-scale="0.5" data-scale-bottom="0.75">

ext/javascript JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:2049,$Easing:$Jease$.$OutQuad},

{$Duration:500,$Delay:80,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Easing:$Jease$.$OutQuad},

{$Duration:1000,x:-0.2,$Delay:40,$Cols:12,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Assembly:260,$Easing:{$Left:$Jease$.$InOutExpo,$Opacity:$Jease$.$InOutQuad},$Opacity:2,$Outside:true,$Round:{$Top:0.5}},

{$Duration:2000,y:-1,$Delay:60,$Cols:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Easing:$Jease$.$OutJump,$Round:{$Top:1.5}},

{$Duration:1200,x:0.2,y:-0.1,$Delay:20,$Cols:8,$Rows:4,$Clip:15,$During:{$Left:[0.3,0.7],$Top:[0.3,0.7]},$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:260,$Easing:{$Left:$Jease$.$InWave,$Top:$Jease$.$InWave,$Clip:$Jease$.$OutQuad},$Round:{$Left:1.3,$Top:2.5}}

];

var jssor\_1\_options = {

$AutoPlay: 1,

$SlideshowOptions: {

$Class: $JssorSlideshowRunner$,

$Transitions: jssor\_1\_SlideshowTransitions,

$TransitionsOrder: 1

},

$ArrowNavigatorOptions: {

$Class: $JssorArrowNavigator$

},

$BulletNavigatorOptions: {

$Class: $JssorBulletNavigator$

}

};

var jssor\_1\_slider = new $JssorSlider$("jssor\_1", jssor\_1\_options);

/\*#region responsive code begin\*/

var MAX\_WIDTH = 1400;

function ScaleSlider() {

var containerElement = jssor\_1\_slider.$Elmt.parentNode;

var containerWidth = containerElement.clientWidth;

if (containerWidth) {

var expectedWidth = Math.min(MAX\_WIDTH || containerWidth, containerWidth);

jssor\_1\_slider.$ScaleWidth(expectedWidth);

}

else {

window.setTimeout(ScaleSlider, 30);

}

}

ScaleSlider();

$Jssor$.$AddEvent(window, "load", ScaleSlider);

$Jssor$.$AddEvent(window, "resize", ScaleSlider);

$Jssor$.$AddEvent(window, "orientationchange", ScaleSlider);

/\*#endregion responsive code end\*/

};

</script>

<style>

/\* jssor slider loading skin spin css \*/

.jssorl-009-spin img {

animation-name: jssorl-009-spin;

animation-duration: 1.6s;

animation-iteration-count: infinite;

animation-timing-function: linear;

}

@keyframes jssorl-009-spin {

from {

transform: rotate(0deg);

}

to {

transform: rotate(360deg);

}

}

.jssorb053 .i {position:absolute;cursor:pointer;}

.jssorb053 .i .b {fill:#fff;fill-opacity:0.5;}

.jssorb053 .i:hover .b {fill-opacity:.7;}

.jssorb053 .iav .b {fill-opacity: 1;}

.jssorb053 .i.idn {opacity:.3;}

.jssora093 {display:block;position:absolute;cursor:pointer;}

.jssora093 .c {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093 .a {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093:hover {opacity:.8;}

.jssora093.jssora093dn {opacity:.6;}

.jssora093.jssora093ds {opacity:.3;pointer-events:none;}

</style>

<div id="jssor\_1" style="position:relative;margin:0 auto;top:0px;left:0px;width:980px;height:380px;overflow:hidden;visibility:hidden;">

<!-- Loading Screen -->

<div data-u="loading" class="jssorl-009-spin" style="position:absolute;top:0px;left:0px;width:100%;height:100%;text-align:center;background-color:rgba(0,0,0,0.7);">

<img style="margin-top:-19px;position:relative;top:50%;width:38px;height:38px;" src="../svg/loading/static-svg/spin.svg" />

</div>

<div data-u="slides" style="cursor:default;position:relative;top:0px;left:0px;width:980px;height:380px;overflow:hidden;">

<div>

<img data-u="image" src="images/pix1.jpg" />

</div>

<div>

<img data-u="image" src="images/pix2.jpg" />

</div>

<div>

<img data-u="image" src="images/pix3.jpg" />

</div>

<div style="background-color:#ff7c28;">

<div style="position:absolute;top:50px;left:50px;width:450px;height:62px;z-index:0;font-size:16px;color:#000000;line-height:24px;text-align:left;padding:5px;box-sizing:border-box;">Photos in this slider are to demostrate jssor slider,<br />

which are not licensed for any other purpose.

</div>

</div>

</div>

<!-- Bullet Navigator -->

<div data-u="navigator" class="jssorb053" style="position:absolute;bottom:12px;right:12px;" data-autocenter="1" data-scale="0.5" data-scale-bottom="0.75">

JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:2049,$Easing:$Jease$.$OutQuad},

{$Duration:500,$Delay:80,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Easing:$Jease$.$OutQuad},

{$Duration:1000,x:-0.2,$Delay:40,$Cols:12,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Assembly:260,$Easing:{$Left:$Jease$.$InOutExpo,$Opacity:$Jease$.$InOutQuad},$Opacity:2,$Outside:true,$Round:{$Top:0.5}},

{$Duration:2000,y:-1,$Delay:60,$Cols:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Easing:$Jease$.$OutJump,$Round:{$Top:1.5}},

{$Duration:1200,x:0.2,y:-0.1,$Delay:20,$Cols:8,$Rows:4,$Clip:15,$During:{$Left:[0.3,0.7],$Top:[0.3,0.7]},$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:260,$Easing:{$Left:$Jease$.$InWave,$Top:$Jease$.$InWave,$Clip:$Jease$.$OutQuad},$Round:{$Left:1.3,$Top:2.5}}

];

var jssor\_1\_options = {

$AutoPlay: 1,

$SlideshowOptions: {

$Class: $JssorSlideshowRunner$,

$Transitions: jssor\_1\_SlideshowTransitions,

$TransitionsOrder: 1

},

$ArrowNavigatorOptions: {

$Class: $JssorArrowNavigator$

},

$BulletNavigatorOptions: {

$Class: $JssorBulletNavigator$

}

};

var jssor\_1\_slider = new $JssorSlider$("jssor\_1", jssor\_1\_options);

/\*#region responsive code begin\*/

var MAX\_WIDTH = 1400;

function ScaleSlider() {

var containerElement = jssor\_1\_slider.$Elmt.parentNode;

var containerWidth = containerElement.clientWidth;

if (containerWidth) {

var expectedWidth = Math.min(MAX\_WIDTH || containerWidth, containerWidth);

jssor\_1\_slider.$ScaleWidth(expectedWidth);

}

else {

window.setTimeout(ScaleSlider, 30);

}

}

ScaleSlider();

$Jssor$.$AddEvent(window, "load", ScaleSlider);

$Jssor$.$AddEvent(window, "resize", ScaleSlider);

$Jssor$.$AddEvent(window, "orientationchange", ScaleSlider);

/\*#endregion responsive code end\*/

};

</script>

<style>

/\* jssor slider loading skin spin css \*/

.jssorl-009-spin img {

animation-name: jssorl-009-spin;

animation-duration: 1.6s;

animation-iteration-count: infinite;

animation-timing-function: linear;

}

@keyframes jssorl-009-spin {

from {

transform: rotate(0deg);

}

to {

transform: rotate(360deg);

}

}

.jssorb053 .i {position:absolute;cursor:pointer;}

.jssorb053 .i .b {fill:#fff;fill-opacity:0.5;}

.jssorb053 .i:hover .b {fill-opacity:.7;}

.jssorb053 .iav .b {fill-opacity: 1;}

.jssorb053 .i.idn {opacity:.3;}

.jssora093 {display:block;position:absolute;cursor:pointer;}

.jssora093 .c {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093 .a {fill:none;stroke:#fff;stroke-width:400;stroke-miterlimit:10;}

.jssora093:hover {opacity:.8;}

.jssora093.jssora093dn {opacity:.6;}

.jssora093.jssora093ds {opacity:.3;pointer-events:none;}

</style>

<div id="jssor\_1" style="position:relative;margin:0 auto;top:0px;left:0px;width:980px;height:380px;overflow:hidden;visibility:hidden;">

<!-- Loading Screen -->

<div data-u="loading" class="jssorl-009-spin" style="position:absolute;top:0px;left:0px;width:100%;height:100%;text-align:center;background-color:rgba(0,0,0,0.7);">

<img style="margin-top:-19px;position:relative;top:50%;width:38px;height:38px;" src="../svg/loading/static-svg/spin.svg" />

</div>

<div data-u="slides" style="cursor:default;position:relative;top:0px;left:0px;width:980px;height:380px;overflow:hidden;">

<div>

<img data-u="image" src="images/pix1.jpg" />

</div>

<div>

<img data-u="image" src="images/pix2.jpg" />

</div>

<div>

<img data-u="image" src="images/pix3.jpg" />

</div>

<div style="background-color:#ff7c28;">

<div style="position:absolute;top:50px;left:50px;width:450px;height:62px;z-index:0;font-size:16px;color:#000000;line-height:24px;text-align:left;padding:5px;box-sizing:border-box;">Photos in this slider are to demostrate jssor slider,<br />

which are not licensed for any other purpose.

</div>

</div>

</div>

<!-- Bullet Navigator -->

<div data-u="navigator" class="jssorb053" style="position:absolute;bottom:12px;right:12px;" data-autocenter="1" data-scale="0.5" data-scale-bottom="0.75">

JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:2049,$Easing:$Jease$.$OutQuad},

{$Duration:500,$Delay:80,$Cols:8,$Rows:4,$Clip:15,$SlideOut:true,$Easing:$Jease$.$OutQuad},

{$Duration:1000,x:-0.2,$Delay:40,$Cols:12,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Assembly:260,$Easing:{$Left:$Jease$.$InOutExpo,$Opacity:$Jease$.$InOutQuad},$Opacity:2,$Outside:true,$Round:{$Top:0.5}},

{$Duration:2000,y:-1,$Delay:60,$Cols:15,$SlideOut:true,$Formation:$JssorSlideshowFormations$.$FormationStraight,$Easing:$Jease$.$OutJump,$Round:{$Top:1.5}},

{$Duration:1200,x:0.2,y:-0.1,$Delay:20,$Cols:8,$Rows:4,$Clip:15,$During:{$Left:[0.3,0.7],$Top:[0.3,0.7]},$Formation:$JssorSlideshowFormations$.$FormationStraightStairs,$Assembly:260,$Easing:{$Left:$Jease$.$InWave,$Top:$Jease$.$InWave,$Clip:$Jease$.$OutQuad},$Round:{$Left:1.3,$Top:2.5}}

];

var jssor\_1\_options = {

$AutoPlay: 1,

$SlideshowOptions: {

$Class: $JssorSlideshowRunner$,

$Transitions: jssor\_1\_SlideshowTransitions,

$TransitionsOrder: 1

},

$ArrowNavigatorOptions: {

$Class: $JssorArrowNavigator$

},

$BulletNavigatorOptions: {

$Class: $JssorBulletNavigator$

}

};

var jssor\_1\_slider = new $JssorSlider$("jssor\_1", jssor\_1\_options);

/\*#region responsive code begin\*/

var MAX\_WIDTH = 1400;

function ScaleSlider() {

var containerElement = jssor\_1\_slider.$Elmt.parentNode;

PYTHON CODE FOR DATASET.CSV

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn import metrics

import matplotlib.pyplot as plt

def confusion\_matrix():

# plot the ROC - confusion matrix graph

data = pd.read\_csv("livestock.csv")

# define the predictor variables and the response variable

X = data[['Cattle\_feeding', 'Cattle\_health', 'Cattle\_score']]

y = data['Cattle']

# split the dataset into training (70%) and testing (30%) sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=1)

# instantiate the model

model = LogisticRegression(solver='lbfgs')

# fit the model using the training data

model.fit(X\_train, y\_train)

# define metrics

y\_pred\_proba = model.predict\_proba(X\_test)[::, 1]

fpr, tpr, \_ = metrics.roc\_curve(y\_test, y\_pred\_proba)

#calculate AUC

auc = metrics.roc\_auc\_score(y\_test, y\_pred\_proba)

return plt, fpr, tpr, auc

def plot\_roc\_curve(plt, fpr, tpr, auc):

# create AUC label

fmt\_auc = float("{:.2f}".format(auc)) # format AUC

plt.plot(fpr, tpr, label="AUC=" + str(fmt\_auc))

# create ROC curve

plt.plot([0, 1], [0, 1], linestyle='-', label='False positive')

plt.plot(fpr, tpr, marker='.', label='True positive')

plt.ylabel('True Positive Rate')

plt.xlabel('False Positive Rate')

plt.legend(loc=4)

plt.show()

# Press the green button in the gutter to run the script.

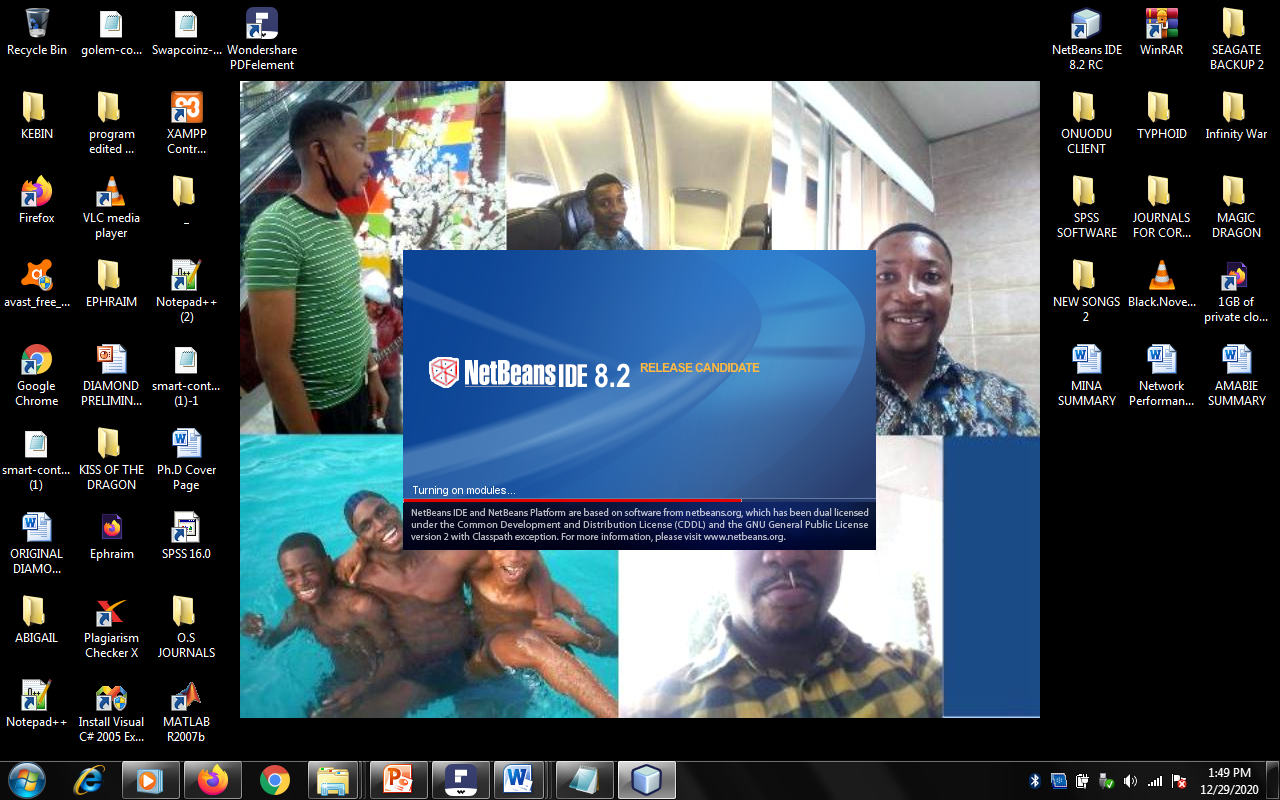
if \_\_name\_\_ == '\_\_main\_\_':

plt, fpr, tpr, auc = confusion\_matrix()

plot\_roc\_curve(plt, fpr, tpr, auc)

**Appendix B**

**Sample Outputs**



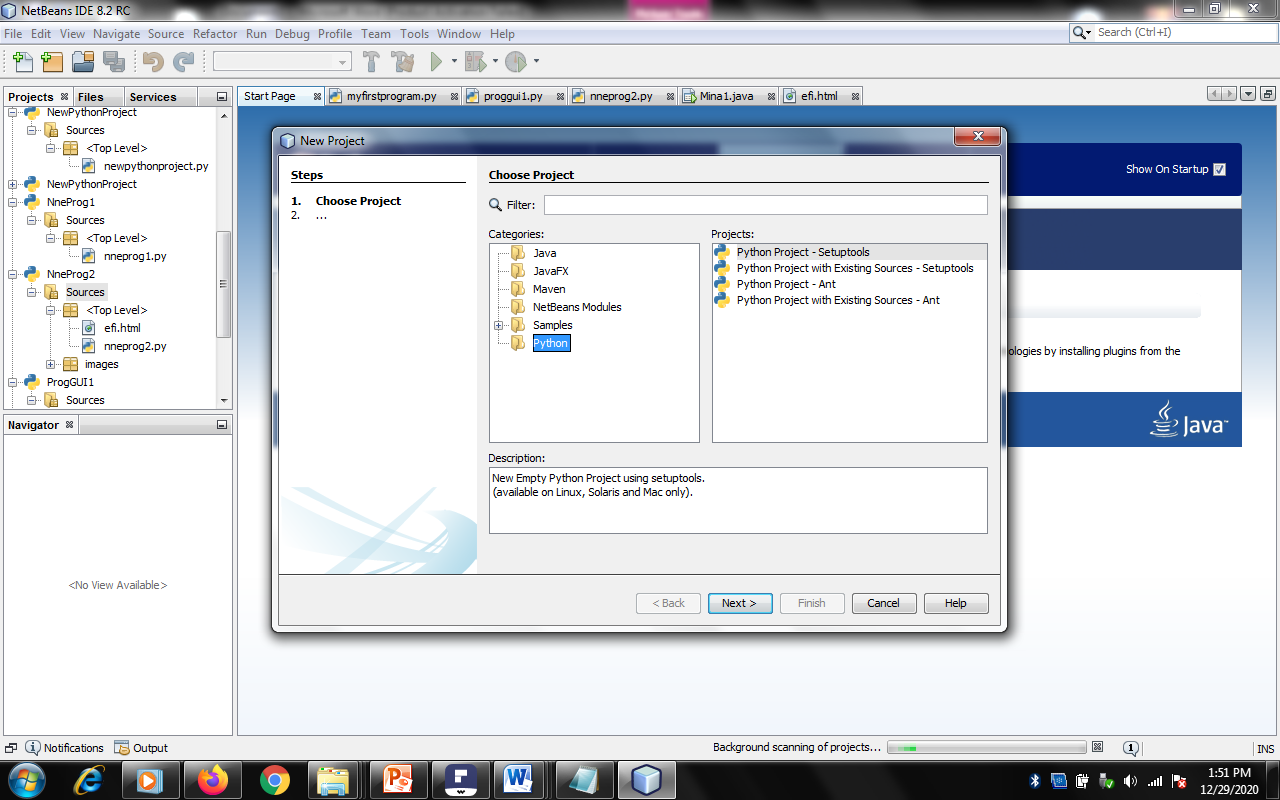


Figure 4.1: Initialization of Python Project in Netbeans 8.2RC IDE



Figure 4.2: Welcome Page

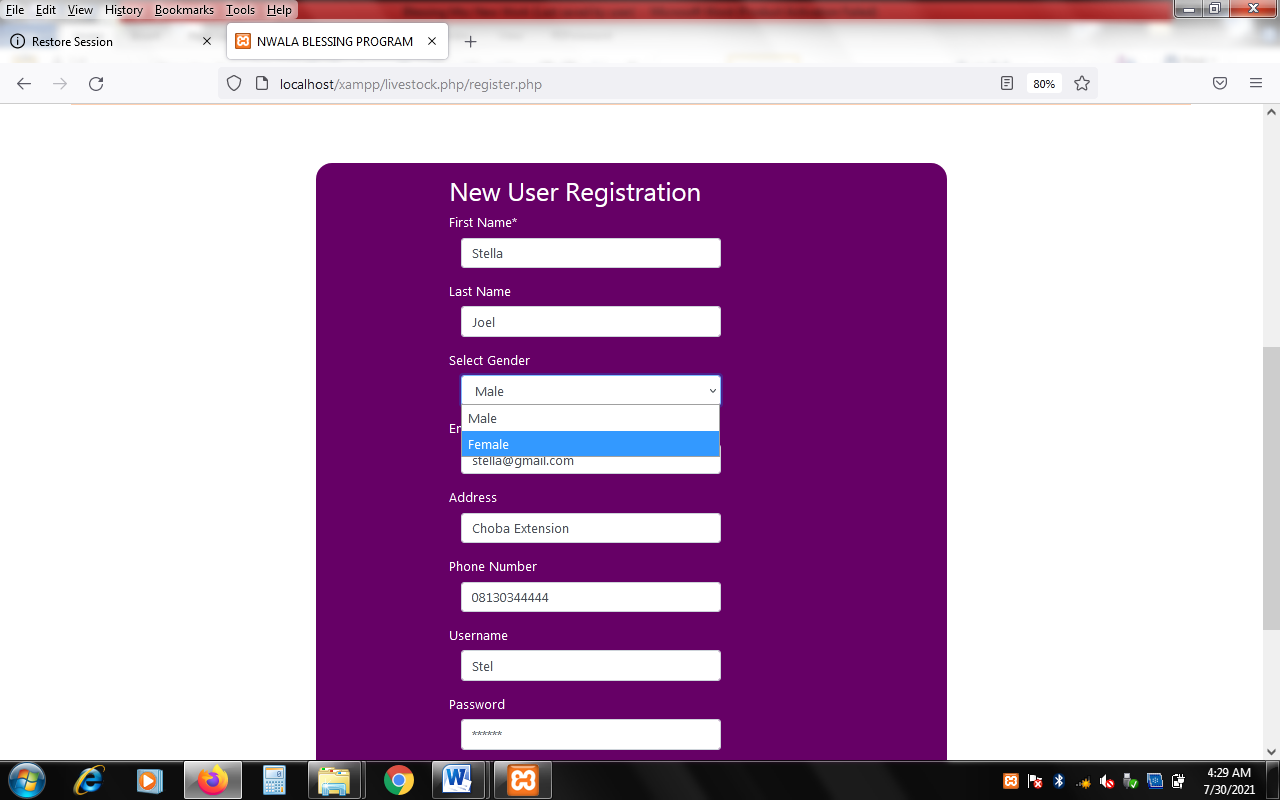


Figure 4.3: Registration Page for new User

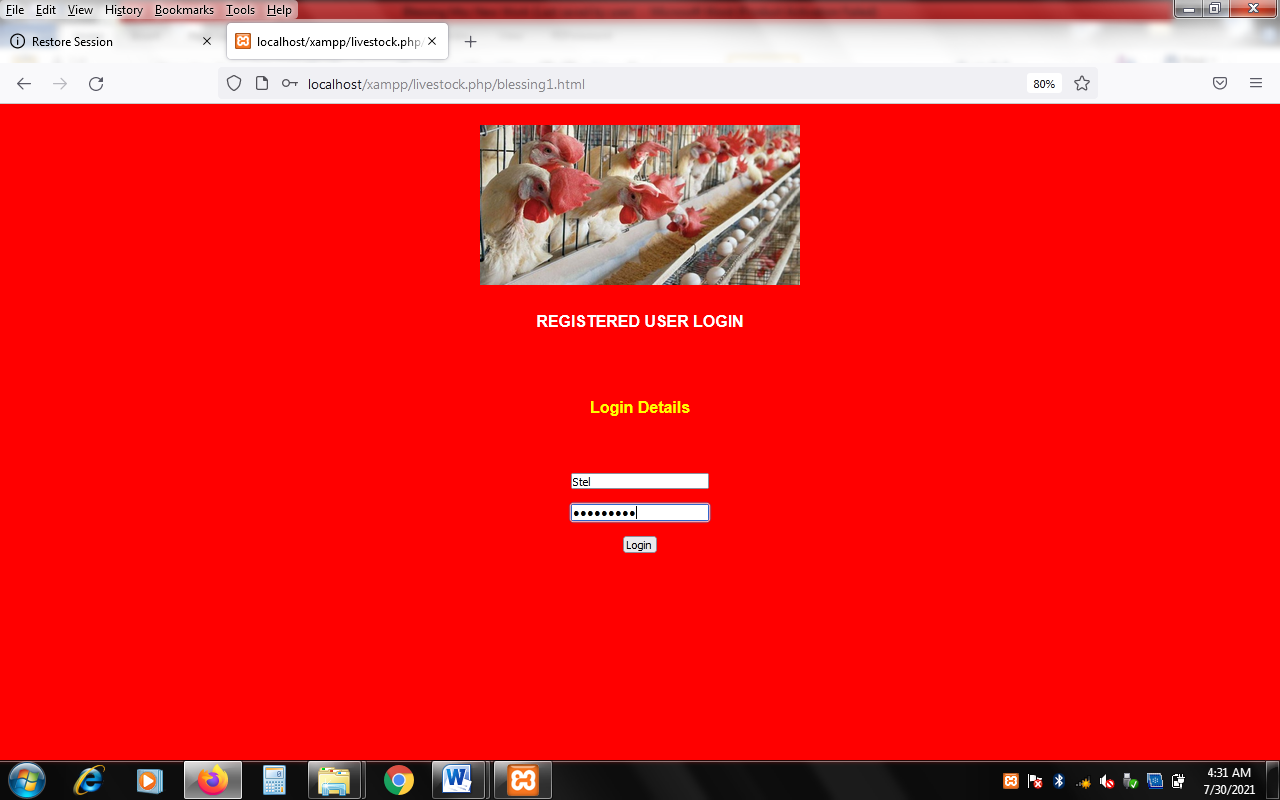


Figure 4.4: Login Page for Registered Users

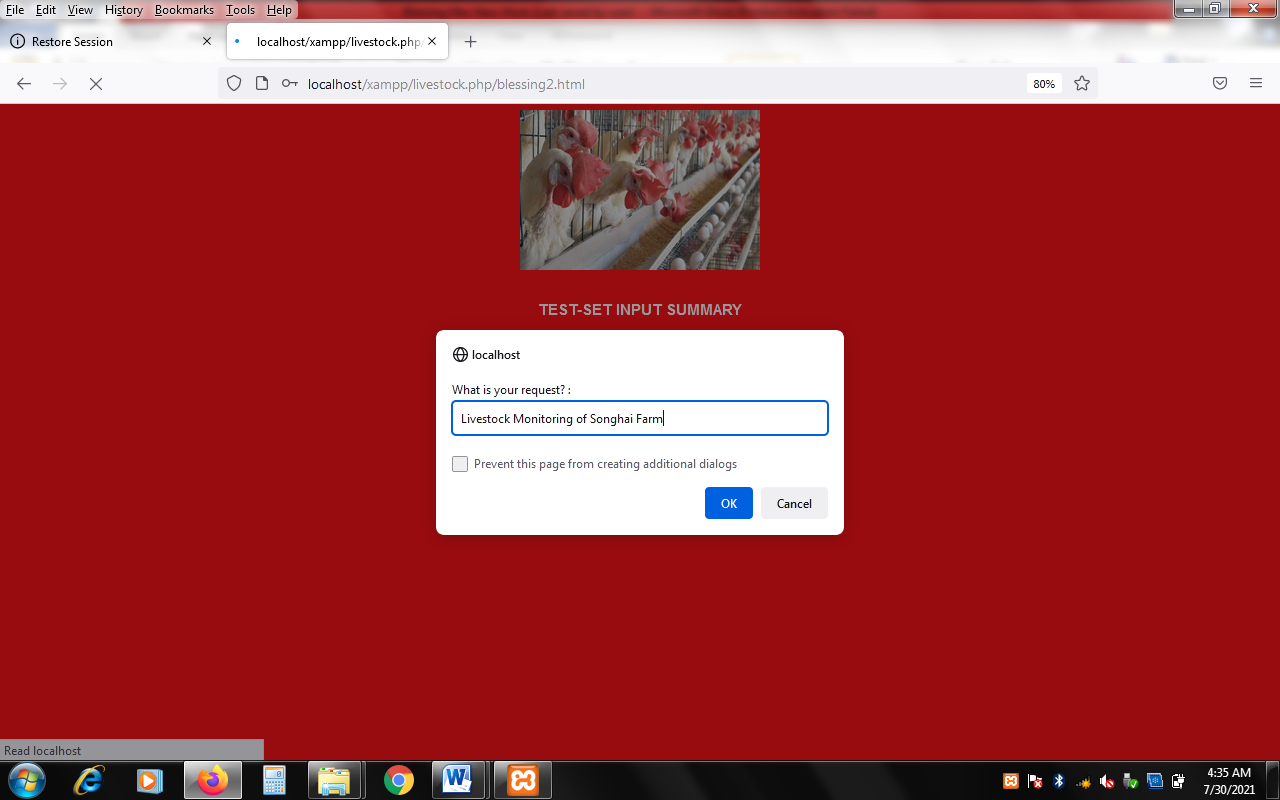


Figure 4.5: Test-Set Input Page

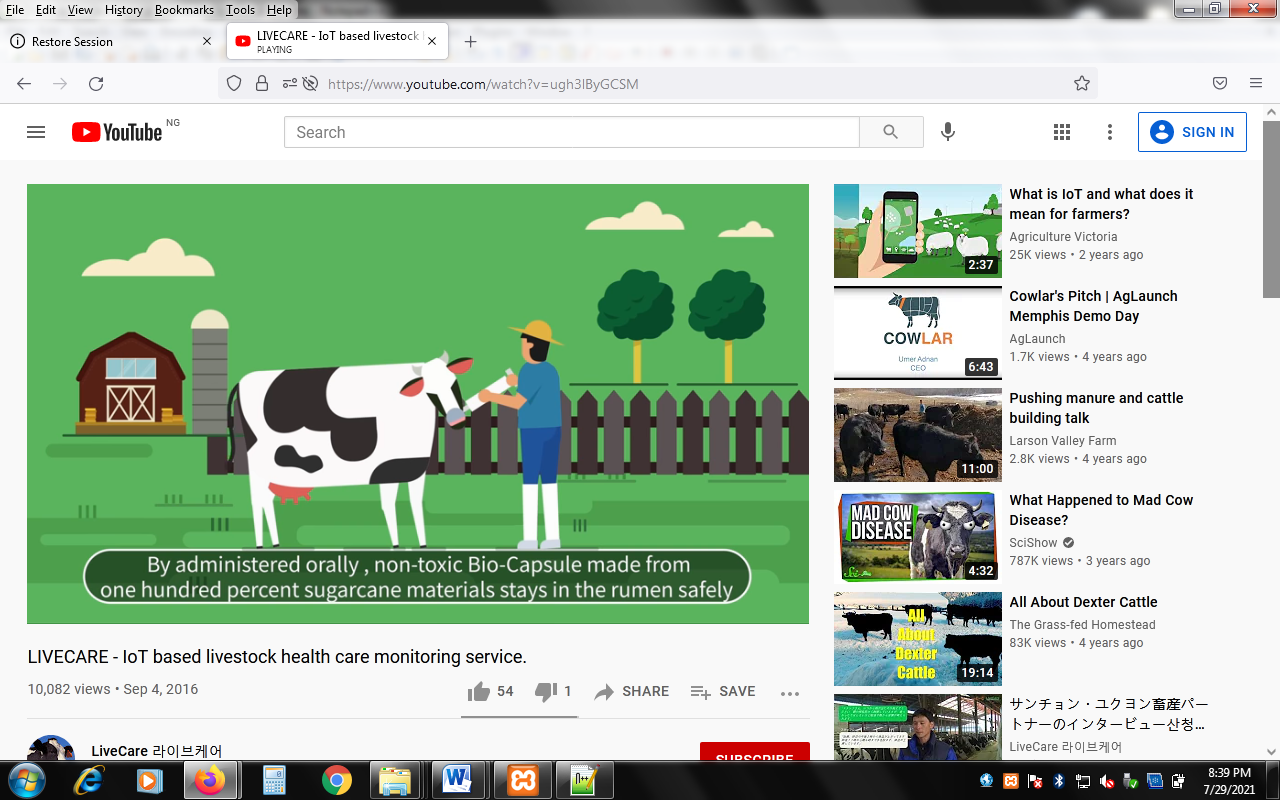


Figure 4.6: Livestock Monitoring and Visualization

**Appendix C**

**Datasets**

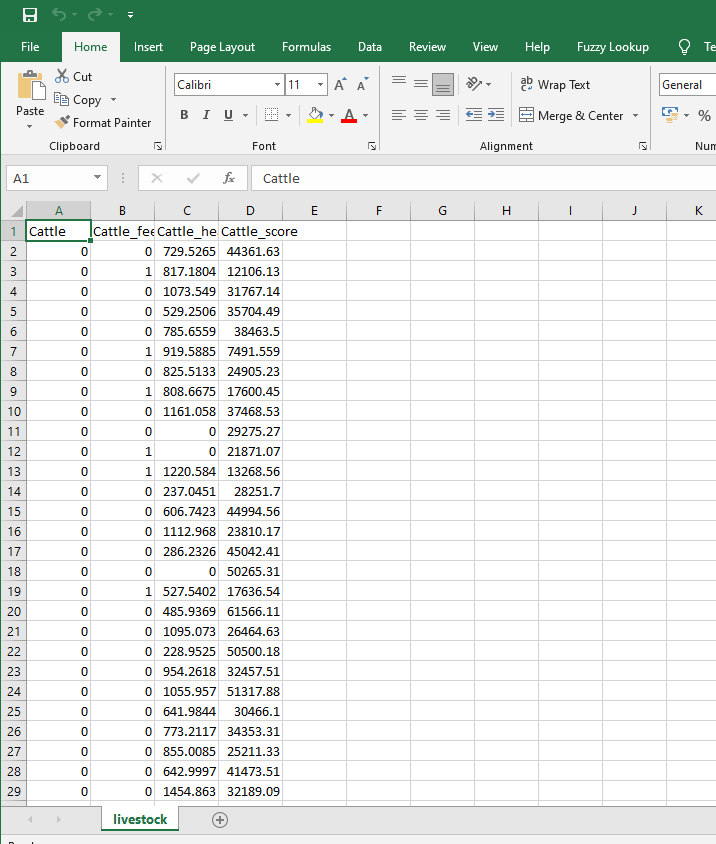


Figure 3.22: Livestock CSV Datasets

(**Source:** https://data.ilri.org/portal/dataset?vocab\_ILRI\_vocsubjects=livestock)



**LIVESTOCK MANAGEMENT SURVEY**

By

**Nwala, Blessing Uchechi**

Research student, Department of Computer Science,

University of Port-Harcourt

**DATE: 8/10/2020**

**FARM LOCATION:**

1. **HOW LONG HAVE YOUR LIVESTOCK FARM BEING IN EXISTENCE?**
2. Day’s b. weeks c. Months d. Years

**2a. HOW MANY CATTLES ARE IN YOUR LIVESTOCK FARM?**

1. 1 – 10 b) 10 – 50 c) 50 – 100

**2b. FROM 2a ABOVE; SPECIFY NUMBER OF BULLS AND COWS**

BULLS =

COWS =

1. **HOW MANY EMPLOYEES WORK IN YOUR LIVESTOCK FARM?**
2. 1-10 b. 10-50 c. 50-100
3. **HOW DO YOU FEED YOUR LIVESTOCK, WHAT TYPE OF ANIMAL FEED IS USED TO FEED YOUR LIVESTOCK?**
4. Natural Feed b. Artificial Feed
5. **HOW LARGE IS YOUR LIVESTOCK FARM?**
6. Small b. Average c. Big d Very big
7. **KINDLY ESTIMATE THE NUMBER BY SPECIE OF LIVESTOCK ON YOUR FARM?**
8. 1-100 b. 100-1000 c. 1000-10000 d. 10000-100000 e. 100000-1000000
9. **WHAT IS THE LEVEL OF LIVE STOCK PRODUCTION INCLUDING BOTH QUANTITY AND QUALITY?**
10. Poor b. Average c. good d. Very good e. Excellent
11. **HAVE YOU IMPLEMENTED THE USE OF SMART MONITORING ON YOUR FARM?**
12. YES B. NO
13. **IF YES WHAT SMART METHOD OR SMART DEVICE DO YOU USE? Please specify.**
14. **WHAT WEATHER CONDITION SUITS YOUR LIVESTOCK IN TERMS OF PRODUCTION AND LIVING?**
15. Cold b. Warm C. Hot
16. **DO YOU OFTEN VACINATE YOUR LIVESTOCK FARM IF YES WHAT TIME OR DURATION?**
17. YES b. NO
18. **HOW DO YOU IDENTIFY TYPE OF DISEASE OUTBREAK ON YOUR FARM? Please specify.**

Regular observation of birds

1. **HOW ARE LIVESTOCKS TREATED DURING SUCH OUTBREAKS? Please Specify.**
2. **WHAT PREVENTIVE METHOD DO YOU USE TO PREVENT DISEASE OUTBREAK ON YOUR FARM? Please specify**.

**Separation of affected cattle’s**

1. **WHAT DETECTIVE METHOD DO YOU USE TO DETECT DISEASE OUTBREAK ON YOUR FARM? Please specify.**

**Answer: Regular human eye observation from the behavior of cattle**

1. **ON A SCALE OF 1-5 HOW DO YOU MONITOR WORKERS BEHAVIOR TOWARDS LIVE STOCK ON YOUR FARM?**
2. 1(poor) b. 2 (average) c. 3(good) d. 4(very good) e. 5(Excellent)
3. **DO YOU HAVE A CONDUCIVE HOME OF SHELTER FOR YOUR LIFE STOCK?**

a). Yes b). No

Table 3.2: Test-set for the new system implementation

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sn** | **Fname** | **Lname** | **Gender** | **Email** | **Address** | **Phone\_No:** | **User**  **name** | **Password** |
| 1. | Mafuzu | Clark | Male | [mc@gmail.com](mailto:mc@gmail.com) | Null | Null | Kun | P01 |
| 2. | Sarah | Williams | Female | [sw@gmail.com](mailto:sw@gmail.com) | Null | Null | Sav | P02 |
| 3. | Olivia | Shane | Female | [os@gmail.com](mailto:os@gmail.com) | Null | Null | Tin | P03 |
| 4. | Hannah | Ethan | Female | [he@gmail.com](mailto:he@gmail.com) | Null | Null | Joh | P04 |
| 5. | Megan | Cameron | Male | [mec@gmail.com](mailto:mec@gmail.com) | Null | Null | JUl | P05 |
| 6. | Chloe | Tim | Male | [ct@gmail.com](mailto:ct@gmail.com) | Null | Null | CT | CT06 |
| 7. | Jess | David | Male | [jd@gmail.com](mailto:jd@gmail.com) | Null | Null | JD | JD07 |
| 8. | Caitlin | Andre | Female | [ca@gmail.com](mailto:ca@gmail.com) | Null | Null | CA | CA08 |
| 9. | Tallulah | Lawrence | Male | [tl@gmail.com](mailto:tl@gmail.com) | Null | Null | TL | TL09 |
| 10. | Ammaarah | Joshua | Female | [aj@gmail.com](mailto:aj@gmail.com) | Null | Null | AJ | AJ10 |
| 11. | Michelle | Johann | Female | [mj@gmail.com](mailto:mj@gmail.com) | Null | Null | MJ | MJ11 |
| 12. | Jenna | Ryan | Female | [jr@gmail.com](mailto:jr@gmail.com) | Null | Null | JR | JR12 |
| 13. | Haajarah | James | Male | [hj@gmail.com](mailto:hj@gmail.com) | Null | Null | HJ | HJ13 |
| 14. | Emma | Matt | Male | [em@gmail.com](mailto:em@gmail.com) | Null | Null | EM | EM14 |
| 15. | Tanja | Xavier | Male | [tx@gmail.com](mailto:tx@gmail.com) | Null | Null | TX | TX15 |
| 16. | Zoe | Calvin | Male | [zc@gmail.com](mailto:zc@gmail.com) | Null | Null | ZC | ZC16 |
| 17. | Maria | Isaac | Female | [mi@gmail.com](mailto:mi@gmail.com) | Null | Null | MI | MI17 |
| 18. | Leah | Armand | Female | [la@gmail.com](mailto:la@gmail.com) | Null | Null | LA | LA18 |
| 19. | Samantha | Luke | Female | [sl@gmail.com](mailto:sl@gmail.com) | Null | Null | SL | SL19 |
| 20. | Ella | Brendan | Female | [eb@gmail.com](mailto:eb@gmail.com) | Null | Null | EB | EB20 |
| 21. | Rachel | Corey | Female | [rc@gmail.com](mailto:rc@gmail.com) | Null | Null | RC | RC21 |
| 22. | Bianca | Thomas | Female | [bt@gmail.com](mailto:bt@gmail.com) | Null | Null | BT | BT22 |
| 23. | Kayla | Bosman | Female | [kb@gmail.com](mailto:kb@gmail.com) | Null | Null | KB | KB23 |
| 24. | Mavita | Bryan | Male | [mb@gmail.com](mailto:mb@gmail.com) | Null | Null | MB | MB24 |
| 25. | Nina | Paci | Male | [np@gmail.com](mailto:np@gmail.com) | Null | Null | NP | NP25 |
| 26. | Danya | Fedinkgoeng | Female | [df@gmail.com](mailto:df@gmail.com) | Null | Null | DF | DF26 |
| 27. | Mailaika | Nkwabi | Male | [mn@gmail.com](mailto:mn@gmail.com) | Null | Null | MN | MN27 |
| 28. | Lily | Errol | Female | [le@gmail.com](mailto:le@gmail.com) | Null | Null | LE | LE28 |
| 29. | Courtney | Treasure | Male | [ct@gmail.com](mailto:ct@gmail.com) | Null | Null | CT | CT29 |
| 30. | Kate | Emershan | Female | [ke@gmail.com](mailto:ke@gmail.com) | Null | Null | KE | KE30 |
| 31. | Sameera | Cowen | Male | [sc@gmail.com](mailto:sc@gmail.com) | Null | Null | SC | SC31 |
| 32. | Laila | Rowan | Female | [lr@gmail.com](mailto:lr@gmail.com) | Null | Null | LR | LR32 |
| 33. | Kendal | Quincy | Female | [kq@gmail.com](mailto:kq@gmail.com) | Null | Null | KQ | KQ33 |
| 34. | Amber | Kreeasen | Male | [ak@gmail.com](mailto:ak@gmail.com) | Null | Null | AK | AK34 |
| 35. | Mia | Sharief | Female | [ms@gmail.com](mailto:ms@gmail.com) | Null | Null | MS | MS35 |
| 36. | Mishka | Sbusiso | Male | [ms@gmail.com](mailto:ms@gmail.com) | Null | Null | MIS | MIS36 |