West African Journal of Open & Flexible Learning
Special Edition: Volume 10, Number 2, 2022

The ODL African Continental Education Strategy: Anchoring AI/Machine Learning on the African Technological Innovation and Investment Table

Professor Gabriel Kabanda
Secretary General
Zimbabwe Academy of Sciences
University of Zimbabwe
gabrielkabanda@gmail.com

Abstract

Through the process of innovation, research produces knowledge and technology that are put to use in the real world. The development of applied scientific technologies can be judged by the number of technological advancements, patents, innovations, research papers that have been published, etc. The purpose of the research was to develop an economic framework and technology solutions for the development of knowledge, innovation, and enterprise on the African continent, using Zimbabwe as a representative example. The study examined the potential applications of cybersecurity and machine learning to the business of producing and disseminating information. The practice of cybersecurity involves a variety of policies, techniques, technologies, and procedures that work together to safeguard the availability, confidentiality, and integrity of computing resources, networks, software programs, and data from intrusion. Cybersecurity is the process of defending systems, networks, and programs from electronic (malicious) attacks. Machine learning (ML) is the process of creating models for the broad relationships among data sets after automatically analyzing massive data sets. Since the Pragmatism paradigm best exemplifies the harmony between knowledge and action, it was chosen as the research philosophy for this study. Focused group discussions served as the main study design while the qualitative component was predominantly used in the knowledge generating component, which was based on an integral research architecture that incorporates descriptive, narrative, theoretical, and experimental survey methodologies. The quantitative dimension investigated machine learning and cybersecurity prototype models using an experiment as a study design. The commercialization of priority projects for

strategic investment included post-harvest technologies, small-scale mining, mineral value addition, and bio-mining, clean water alternatives, mining waste-derived tile technologies, ICT innovations in machine learning and cybersecurity, and defense technologies. To direct the installation of upcoming cybersecurity systems in Africa, a Bayesian Network model for cybersecurity was developed. The research developed an effective network intrusion detection system using the KDDCup 1999 intrusion detection benchmark dataset. The sample included 494,020 instances of primary data with 42 variables that were analyzed mostly using the SNORT open source program and other Bayesian Network-supporting platforms. The most effective ML algorithms were considered while creating a Bayesian Network model.

Keywords: *Knowledge generation, innovation, sustainable development, economic framework, Cybersecurity, Artificial Intelligence, Machine Learning.*

http://orcid.org/0000-0001-6699-080X

1.0 INTRODUCTION

1.1 Background

Through the process of innovation, research produces knowledge and technology that are put to use in the real world. In order to create new ways of knowing, disciplinary traditions' value-adding material and methods are combined in a process known as knowledge co-creation. Knowledge co-creation, as a management initiative, results in the blending of ideas and the harmonization of several parties to jointly develop an outcome that is valued by both sides. Innovation is the successful application of a concept, updating and expanding the selection of goods and services, establishing new processes for manufacturing, supply, and distribution, introducing changes to management, and reorganizing the way that people operate. Innovation is the successful application of a concept, updating and expanding the selection of goods and services, establishing new processes for manufacturing, supply, and distribution, introducing changes to management, and reorganizing the way that people operate. Innovation is also the process of developing and utilizing combinations of information from various sources to have an impact on progress and to

generate revenue from inventions. Technology comprises objects, knowledge, actions, processes, and an associated socio-technical system (Kabanda G., 2013). The development of applied scientific technologies can be judged by the number of technological advancements, patents, innovations, research papers that have been published, etc.

The Zimbabwe National Vision 2030 is ***"Towards a Prosperous and Empowered Upper Middle-Income Society by 2030, with Job Opportunities and a High Quality of Life for its Citizens".*** The ultimate objective of Vision 2030 is to raise Zimbabwe's per capita Gross National Income from its present level of US$1,440 to over US$5,000 in real terms by 2030, making it an upper middle-income economy. The National Development Strategy was developed based on the National Vision for Zimbabwe, expressed as Vision 2030, and is currently being executed as the two subsequent Five-Year National Development Strategies, NDS1 (2021-2025) and NDS 2. (2026-2030).

The Internet of Things (IoT) era produces enormous volumes of data that are gathered from numerous heterogeneous sources, such as mobile devices, sensors, and social media. Using a hybrid cybersecurity architecture that combines machine learning and artificial intelligence techniques, cloud computing environments may be protected from IoT cyber-attacks. In order to prevent attacks, cybersecurity combines the confidentiality, integrity, and availability of computing resources, networks, software, and data into a coherent set of policies, technologies, processes, and strategies (Berman, D.S., et al, 2019). The goal of cybersecurity is to preserve and safeguard computer information systems from cyberattacks or illegal access. It consists of a number of technologies, processes, and operations (Sarker, I.H., et al, 2020). The two main cybersecurity applications, intrusion detection and malware detection, have forced a drastic change in cybersecurity's technology and operations in order to detect and eradicate cyber threats and reduce costs associated with compromised systems, networks, and data (Sarker, I.H., et al, 2020).
By teaching computers to think and behave like people, artificial intelligence (AI) imitates human intellect in technology (Nielsen, R., 2015). Machine Learning (ML) is a subset of AI where machines are

given learning instructions. In essence, machine learning (ML) is the automatic analysis of massive amounts of data and the creation of models to determine the correlations between the data. Empirical data is a necessary input for ML algorithms, which then learn from this data. The three classes of ML, according to Truong, T.C., et al (2020), are as follows:

i. *Supervised learning* occurs when the algorithms are given training examples in the form of inputs labeled with the anticipated outcomes. Supervised learning is a machine learning job that infers a function from labelled training data made up of a collection of training examples. It involves learning a function that maps an input to an output using example input-output pairs.
ii. *Unsupervised learning* involves algorithms that learn unsupervisedly when they are given unlabeled inputs. In contrast to supervised learning, where there are correct answers and a teacher, the algorithms are left to find and display the interesting structure in the data on their own.
iii. *Reinforcement learning* takes place when action sequences, observations, and rewards are used as inputs. The field of machine learning known as reinforcement learning looks at how software agents should behave in a given environment to maximize a theoretical concept of cumulative reward.

Machine learning is a big data analytics technique that primarily entails programming the creation of analytical models (Napanda, K., et al, 2015). The development of big data analytics as a field of techniques for data mining and analysis most suitable for enormous datasets beyond the capacity of conventional data-processing approaches (Nielsen, R., 2015). Big Data emerged as a result of the inability of standard relational database systems to handle the unstructured data produced by businesses, social media, and other data-generating sources (Mazumdar, S., and Wnga, J., 2018). Cyberattacks have gotten more sophisticated and complex in an era of transformation and expansion in the Internet of Things (IoT), cloud computing services, and big data (Wilson, B.M.R., et al., 2015). As a result, cybersecurity events are becoming more challenging or impossible to detect using traditional detection systems (Hashem, I.A.T., et al, 2015; Siti,N.M., et

al, 2017). Regarding the provision of security dimensions in network traffic management, web transaction access patterns, network server configuration, data sources for the network, and user identity and authentication information, Big Data Analytics (BDA) is rich in capabilities. The fields of security management, identity and access management, fraud prevention, governance, risk, and compliance have seen a significant revolution as a result of these operations.

The Support Vector Machine is a supervised machine learning technique that can be applied to classification and regression problems (SVM). The k-nearest neighbors (KNN) technique is the simplest and easiest supervised machine learning approach that can resolve both classification and regression issues. When determining the best handover solutions in heterogeneous networks made up of several cell types, both the SVM and KNN can be used. Machine learning algorithms have the capacity to take advantage of the user context learned given a collection of contextual input cues. Regression models, K-nearest neighbors, Support Vector Machines, and Bayesian learning are among the supervised learning techniques (Thomas, E.M., et al, 2013).

Monitoring activity on networked computer systems and conducting analysis of potential breaches or violations of different computer security policies constitute intrusion detection. The escalating usage of the internet and the dangers it poses has given rise to network intrusion detection systems (NIDS). A sort of computer program called NIDS monitors system activity to spot behavior that violates security policy and can distinguish between malicious and legitimate network users (Bringas, P.B., and Santos, I., 2010, p.229). Misuse network detectors and anomaly detectors are the two types of NIDS. Systems for detecting misuse thoroughly inspect every incoming network traffic and find any sequence that appears in the database. On the other hand, anomaly detection systems concentrate on identifying novel, undiscovered risks (Bringas, P.B., and Santos, I., 2010, p.229). ML in particular is strongly reliant on AI paradigms for anomaly identification. The best tool for achieving this integration of misuse network detectors and anomaly detectors is a set of Bayesian networks.

In order to represent graphical probabilistic models for multivariate analysis, Bayesian Networks (BNs), which are directed acyclic graphs with an associated probability distribution function, are utilized (Bringas, P.B., and Santos, I., 2010, p.231). A Bayesian network is simply defined by Boudali, H., and Dugan, J.B. (2006, p.86) as a directed acyclic graph with nodes and arcs, where the nodes stand in for random variables (RV) and the directed arcs between pairs of nodes signify dependencies between the RV. The probability function also shows how strong these connections are in the graph. Formally, let a Bayesian Network B be defined as a pair, B = (D, P), where D is a directed acyclic graph; P = $\{p(x_1|\Psi_2), ..., p(x_n|\Psi_n)\}$ is the set composed of n conditional probability functions (one for each variable); and Ψi is the set of parent nodes of the node X_i in D. The set P is defined as the joint probability density function (Bringas, P.B., and Santos, I., 2010, p.232).

$$p(x) = \prod_{i=1}^{n} p(x_i \mid \Psi_i)$$

A Bayesian network G is a probabilistic graphical model that encodes a joint probability distribution over a set of variables X = $\{X_1, X_2, ..., X_n\}$ based on conditional independencies. This is also a directed acyclic graph (DAG) where each node represents a random variable and an edge denotes a direct probabilistic dependency between the two connected nodes (Xiao, L., 2016, p.10).

The Bayesian network accurately represents the joint probability distribution as

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid PaG(X_i))$$

where $PaG(X_i)$ denotes the set of parent nodes of X_i in G, and $p(X_i|PaG(X_i))$ specifies the conditional probability distribution (CPD) of X_i given $PaG(X_i)$ (Xiao, L., 2016, p.10).

The capacity of Bayesian Networks to calculate the likelihood that a particular hypothesis is correct from a historical dataset is one of its greatest strengths. The arguments for using Bayesian networks for this kind of research, according to Margaritis, D. (2003, p. 2), are that they:

1. Are graphical models, capable of presenting relationships clearly and intuitively.
2. They can reflect cause-and-effect interactions since they are directed.
3. Able to deal with uncertainty.

1.2 Statement of the Problem

Zimbabwe's limited ability to industrialize and modernize presents a challenge to sustainable development. Breaking down silos, using synergies, and forging clever collaborations within the National Science, Technology, and Innovation System are necessary to generate wealth and construct an innovation-led knowledge economy (NSTIS). The budgetary space available to the government of Zimbabwe is constrained, making it difficult for it to raise money to undertake socio-economic programmes.

Due to glaring shortcomings against external threats, firewall protection for computer systems and networks in Information Communication Technologies (ICTs) has shown to be insufficient. The reality is that the majority of network-centric cyberattacks are perpetrated by intelligent agents, making it necessary to combat them using intelligent semi-autonomous agents that can recognize, assess, and react to cyberattacks (Kabanda, G., 2021). The rapid advancement of computing and digital technologies has made it necessary for most organizations to update their cyberdefense strategy (Kabanda, G., 2021). In order to effectively detect cyber threats in real time, security network managers must be more agile, flexible, and provide strong cyber defense systems. Its of paramount importance to explore the opportunities of Machine Learning (ML) and Big Data Analytics (BDA) paradigms for use in Cybersecurity.

1.3 Purpose or Aim

The purpose of the research was to formulate an economic framework and develop technological solutions for Zimbabwe with respect to knowledge generation, innovation and enterprise development. The

ultimate goal was to develop, exploit, and commercialize at least two priority applied scientific inventions in 100 days, after which potential in machine learning and cybersecurity would be investigated.

1.4 Main Research Question

In the light of technical advancements like machine learning and cybersecurity, how do we create an economic framework and technological solutions for Zimbabwe's sustainable development?

1.5 Research Questions

a) How does Zimbabwe's National Science, Technology, and Innovation System (NSTIS) generate knowledge?
b) How do you utilize and market cutting-edge technology solutions?
c) What constitutes successful knowledge creation and business development in Zimbabwe's NSTIS?
d) How are the Big Data Analytics and Machine Learning paradigms employed in cybersecurity to provide secure ICTs?
e) How can a Bayesian network model be created that can manage the complexity of cybersecurity?

1.6 Rationale and Justification for the Research

According to the African Union's Agenda 2063, the vision of the African Union is “An integrated, prosperous and peaceful Africa, an Africa driven and managed by its own citizens and representing a dynamic force in the international arena”. The mission of the Science, Technology and Innovation Strategy for Africa (STISA-2024) is to "accelerate Africa 's transition to an innovation-led, knowledge-based economy". Table 1 below lists the six main socioeconomic goals that will be addressed by the STISA-2024 research and innovation priority areas for Africa. Of these, priority areas 3 (communication - physical and intellectual mobility) and 6 are of particular relevance (wealth creation).

Table 1: STISA 2024 Priority Areas

PRIORITIES	RESEARCH AND/OR INNOVATION AREAS

1. Eradicate hunger and ensure food and nutrition security	Agriculture/ Agronomy in terms of cultivation technique, seeds, soil and climate Industrial chain in terms of conservation and/or transformation and distribution infrastructure techniques
2. Prevent and Control Diseases and ensure Well-Being	Better understanding of endemic diseases - HIV/AIDS, Mal aria Hemoglobinopathe Maternal and Child Health Traditional Medicine
3. Communicatio n (Physical and Inte llectual Mobility)	Physical communication in terms of land, air, river and maritme routes equipment and infrastructure and energy Promoting local materials Intellectual communications in terms of ICT
4. Protect our Space	Environmental protection including climate change studies Biodiversity and Atmospheric Physics Space technologies, maritime and sub-maritime exploration Knowledge of the water cycle and river systems as well as river basin management
5. Live Together - Build the Society	Citizenship, History and Shared Values Pan Africanism and Regional Integration Governance and Democracy, City Management and Mobilit y Urban Hydrology and Hydraulics Urban Waste Management
6. Create Wealth	Education and Human Resource Development Exploitation and management of mineral resources, forests, aquatics, marines, etc. Management of water resources

The National Innovation System (NIS) is a collection of institutions that work together to propel a country's creative performance. Any nation's innovation system frequently consists of institutions, rules, laws, and guidelines for the production, sharing, preserving, and application of knowledge. Excellence, creativity, and leadership are the key success factors in each of these endeavors. The graphic on Figure 2 serves as an illustration of Zimbabwe's National Science, Technology, and Innovation System. The three components of innovation—creativity, entrepreneurship and commercialization, and

dissemination and adaptation—do, however, need to be carefully balanced.

Poverty reduction has been a top priority for the Zimbabwean government, as was the main objective of the publicly released *"Zimbabwe Agenda for Sustainable Socio-Economic Transformation" for the period from October 2013 to December 2018* (Zim-Asset). The National Development Strategy is now being implemented by the Government of Zimbabwe through the Ministry of Finance and Economic Planning of Zimbabwe, which is doing so through two subsequent Five-Year National Development Strategies: NDS1 (2021-2025) and NDS 2. (2026-2030).

2.0 REVIEW OF LITERATURE

2.1 The Sustainable Development Goals (SDGs) Context

A group of eminent researchers from the Zimbabwe Academy of Sciences were selected by the Research Council of Zimbabwe (RCZ) to work on an innovative research project with the goal of "generating, utilizing, and commercializing at least two priority applied scientific technologies within 100 days." The Sustainable Development Goals (SDGs), the priority areas of the African Science, Technology and Innovation Strategy for Africa (STISA 2024), Vision 2030, and the national research priority areas all inform the Knowledge Generation research priority areas for Zimbabwe. The United Nations published the SDGs where SDG Goal 9 is about the need to *"build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation"* (*http://www.un.org/sustainabledevelopment/sustainable-development-goals/).* Zimbabwe as a country prioritised goals 2,3,4,5,6,7,8,9,13 and 17. However, Zimbabwe as a nation must not be viewed as just gravitating from one developmental guideline to another without making any meaningful progress during each dispensation.

The vision for the Zimbabwe Academy of Sciences (ZAS) is "*The Zimbabwe Academy of Sciences seeks to be the leading catalyst for knowledge-sharing, innovative solutions, evidence-based policy formulation and advisory services in Zimbabwe, Africa and beyond".* In accordance with its mission statement, ZAS's purpose is to *"monitor*

the environment, identify issues and opportunities, and provide and communicate the best evidence-based solutions that benefit society for sustainable development by mobilizing the science community and other resources through smart partnerships with government, academia, the private sector, development partners, and civil society." However, the revitalized Mission must be self-renewing, task-oriented, relevant, agile, flexible, and consistent in order to find creative solutions to Zimbabwe's problems and strategically propel Zimbabwe to become a major world power. The ZAS guiding philosophy is about mutual respect and quality, stated clearly as *"Mutual respect and equality is important because my humanity is bound up with yours".* The fundamental principles of innovation, integrity, professionalism, dependability, institutional independence, respect, and ethics serve as a support for this. In accordance with the national research priorities and significant key projects of national import, ZAS aspires to provide national leadership on scientific initiatives and innovations in key fields such as heritage studies, water and sanitation, climate change, sustainable environmental management, national security, etc. The following are the top national research categories:

1. Social Sciences and Humanities
2. Sustainable Environmental and Resource Management
3. Promoting and Maintaining Good Health
4. National Security

The new priority areas that require special attention include the following:

1. Natural and Cultural Heritage
2. Indigenous Knowledge Systems
3. Post-harvest technologies
4. Rural transformation
5. Small scale mining/mineral value addition/bio mining
6. Clean water alternatives
7. Tiles technologies from mining waste
8. Cyber security systems
9. Defence technologies (double use technologies (drones)

Figure 1 below shows how the 17 Sustainable Development Goals (SDGs) describe the main development issues that humanity must

overcome in order to ensure a sustainable, peaceful, prosperous, and equitable way of life. Zimbabwe prioritised Goals 2, 3, 5, 6, 7, 8, 9, 13, and 17 priority. The 17 Sustainable Development Goals (SDGs) are the cornerstone of the 2030 Agenda and outline the greatest development issues facing humanity in order to ensure a sustainable, peaceful, prosperous, and equitable way of life. Peacebuilding and promoting sustainable development are essential for humanity. Due to the restricted ability of emerging countries to industrialize and modernize, there is an issue with sustainable development.

The African Union has a clearly defined Mission on Science, Technology and Innovation (STI), which was considered by Zimbabwe as an African country.The African Union, through its Agenda 2063, desires a prosperous and peaceful Africa. The Science, Technology and Innovation Strategy for Africa ***(STISA-2024)*** is to ***"accelerate Africa 's transition to an innovation-led, knowledge-based economy".*** Breaking down silos, leveraging synergies, and forming clever collaborations in the National Science, Technology, and Innovation System are necessary at the national level to generate wealth and construct an innovation-led knowledge economy (NSTIS).

Figure 1: The United Nations Sustainable Development Goals

ZAS oversees the creation of scientific knowledge in Zimbabwe, which involves a number of different parties, including the Zimbabwe Academy of Sciences, sectoral research councils, universities and colleges, statistical agencies, standards measurement bodies, public laboratories, research centers, private laboratories, intellectual property agencies, custodians of indigenous knowledge, heritage statutory bodies, etc. The Zimbabwe Academy of Sciences collaborates with a number of national organizations to achieve its goals and objectives. These institutions include the government, academia, research, and legislators, among others.

The basic driving force behind economic growth is technological change, where the main catalyst is investment on research and development. Israel has made significant investments in research and development, as seen in Table 2 below, in comparison to Zimbabwe (Kabanda G., 2013).

Table 2: Effects of R&D to economic performance

	Zimbabwe	Israel
Population	14,627,000	7,700,000
Size	200,000 SQ km	20,000 SQ km
GDP - Per Capita	***$1,530***	***$31,004***
Infant Mortality Rate	***79*** dead per 1000 live births	***4 dead*** per 1000 live births
Life Expectancy	***50***	***81***

Israel, which has half the population of Zimbabwe and ten times less land than Zimbabwe, has made significant investments in research and development (R & D) over the years, and as a result, its economy has advanced to join the ranks of developed nations. Israel now boasts a GDP per capita of $31,004 and a life expectancy of 81 years, compared to Zimbabwe's $1,530 and 51 years, respectively. Due to significant R&D expenditures, Israel, whose desert-covered half of the country, currently exports more agricultural and citrus products than both Zimbabwe and South Africa combined. How does Zimbabwe improve its average life expectancy of 50 years to the life expectancy of Israel of 81 years, or improve its GDP per capita from just $1,530 to $31,004 like Israel? High GDP per capita is demonstrated in Southern Africa (SADC Region) by Seychelles, which has a GDP of $16,434, Mauritius, which has a GPD per capita of $11,228 and Botswana, which has a $8,258. Namibia's GDP per capita is $6,013 while South Africa's is $6,354.

The GDP per capita analysis for Southern Africa as of August 2021 is shown on the table 3 below.

Table 3: SADC Regional GDP per capita - August 2021

Country	Population	Annual GDP (US$)	GDP per capita (US$)
Angola	30,809,762	105, 902M	3,437
Botswana	2,254,126	18,615M	8,258
DRC	84,068,091	47,099M	560
Lesotho	2,108,132	2,739M	1,299
Madagascar	26,262,368	13,853M	528
Malawi	18,143,315	7,065M	389
Mauritius	1,266,000	14,210M	11,228
Mozambique	29,495,962	14,396M	488
Namibia	2,414,000	14,513M	6,013
Seychelles	96,762	1,590M	16,434
South Africa	57,939,000	368,135M	6,354
Swaziland	1,136,191	4,711M	4,146
Tanzania	56,318,348	56,852M	1,009
Zambia	17,351,822	26,720M	1,540
Zimbabwe	14,439,018	20,401M	1,530

The Zimbabwe Academy of Sciences (ZAS) was founded in October 2004 as a result of research work carried out by the Research Council of Zimbabwe. Its goals are to provide the government and the country at large with independent, evidence-based advice on how to address national challenges using scientific knowledge and cutting-edge expertise as well as to recognize, honor, and perpetuate the accomplishments of those Fellows of ZAS who have significantly contributed to the development of science.

The Knowledge Generation work for the NSTIS of Zimbabwe is guided by Goal 9 of the globally defined 17 Sustainable Development Goals (SDGs), which largely inform some of the national developmental programmes, are listed in Table 4 below. Zimbabwe as a country prioritised goals 2,3,4,5,6,7,8,9,13 and 17.

Table 4: The 17 Sustainable Development Goals (SDGs)
(*http://www.un.org/sustainabledevelopment/sustainable-development-goals/)*

Goal 1	End poverty in all its forms everywhere
Goal 2	End hunger, achieve food security and improved nutrition and promote sustainable agriculture
Goal 3	Ensure healthy lives and promote well-being for all at all ages
Goal 4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
Goal 5	Achieve gender equality and empower all women and girls
Goal 6	Ensure availability and sustainable management of water and sanitation for all
Goal 7	Ensure access to affordable, reliable, sustainable and modern energy for all
Goal 8	Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
Goal 9	***Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation***
Goal 10	Reduce inequality within and among countries
Goal 11	Make cities and human settlements inclusive, safe, resilient and sustainable
Goal 12	Ensure sustainable consumption and production patterns
Goal 13	Take urgent action to combat climate change and its impacts*
Goal 14	Conserve and sustainably use the oceans, seas and marine resources for sustainable development
Goal 15	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss
Goal 16	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
Goal 17	Strengthen the means of implementation and revitalize the global partnership for sustainable development

The ICT revolution will require money for the necessary equipment and infrastructure, as well as significant investments in labour (Kabanda G., 2008). From an ICT perspective, the revolutionary technical development or productivity levels are related to labour and

capital by the Cobb-Douglas production function, which has the following form:

$$\mathbf{Q = A K^a L^b}$$

is used for the analysis of technological progress and attended economic growth, where A, a and b are empirical parameters**.**

- K = capital input (very meaningful mounts)
- L = labour input (high technical competence)

Through an investment in technology, production capacity can be doubled several times over. Economic growth is mostly driven by technological advancement. Technological development is endogenous because it results from deliberate actions taken by economic agents in response to financial incentives. High-skilled labor is complementary to capital and low-skilled labor because it accelerates technical advancements and their dispersion. The issue of "Brain Drain" arises from a nation's ongoing loss of highly skilled workers. The graphic below, which links the output per worker to the capital per worker in how it connects to the investment per worker and the output per worker, serves as an example of an endogenous neoclassical economic growth model.

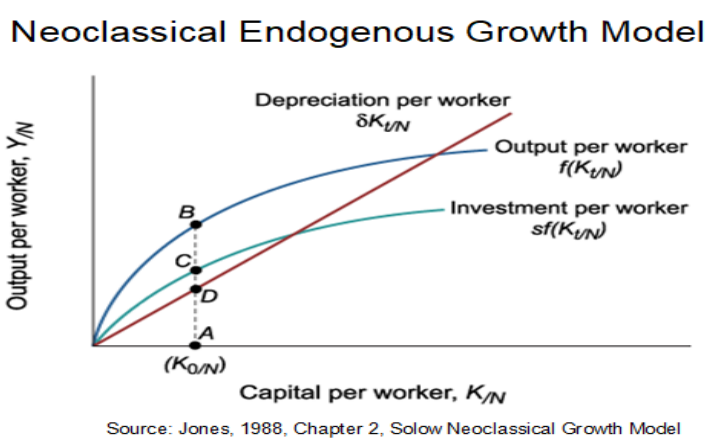


Figure 2: Neoclassical Endogenous Growth Model

2.2 Classical Machine Learning (CML)

Machine Learning (ML) is a field in artificial intelligence where computers learn like people. As indicated in Table 5, we list and briefly explain the most popular classical machine learning algorithms. The following are some examples of common classical ML algorithms shown on Table 5.

Table 5: Classical Machine Learning Algorithms

Classical Machine Learning Algorithm	Explanation
i. Logistic Regression (LR)	Logistic regression is similar to linear regression in concept and was developed by Sit, N.M., et al (2017), but it prevents misclassification that could happen in linear regression. Findings from logistic regression are essentially either "0" or "1," unlike results from linear regression. The quantity of training data has a significant impact on the effectiveness of logistic regression.
ii. Naive Bayes (NB)	The Naive Bayes (NB) classifier is founded on the Bayes theorem, which presupposes feature independence. The Naive Bayes classifier escapes the dimensionality plague thanks to the independence assumptions.
iii. Decision Tree (DT)	A decision tree has a structure similar to a flowchart, with the root node at the top and each internal node designating an aspect of the information. Due to the fact that even a small change in the information would alter the tree's structure, the algorithm may be biased and ultimately unstable.
iv. Nearest Neighbor (KNN)	K-Nearest Neighbor (KNN) is a non-parametric method that employs similarity measures in relation to classifiers that use distance functions rather than news cases. KNN is computationally expensive since it saves all of the training data in a bigger amount of memory.
v. Ada Boost (AB)	A method for improving the efficiency of straightforward learning algorithms used for classification is the Ada Boost (AB) algorithm. Ada Boost combines a number of weak classifiers to create a strong classifier. It is a quick classifier that can also function as a feature learner simultaneously. This could be helpful for tasks involving the analysis of unbalanced data.
vi. Random Forest (RF)	As an ensemble technique, random forest (RF)

	generates a decision tree from a subset of observations and variables. A single decision tree cannot predict as well as the Random Forest. It builds a number of minimally correlated decision trees using the bagging principle.
vii. Support Vector Machine (SVM)	The supervised machine learning technology known as the Support Vector Machine (SVM) can be used to handle classification and regression issues. SVM is a linear classifier with a hyperplane as the classifier. The training set is separated by the widest possible margin. Support vectors are the places close to the dividing hype plane that dictate where the hyper plane will be.

2.3 Modern Machine Learning

Deep learning has the capacity to intuitively learn the best feature representation from the most basic inputs.

2.2.1 Deep Neural Network (DNN)

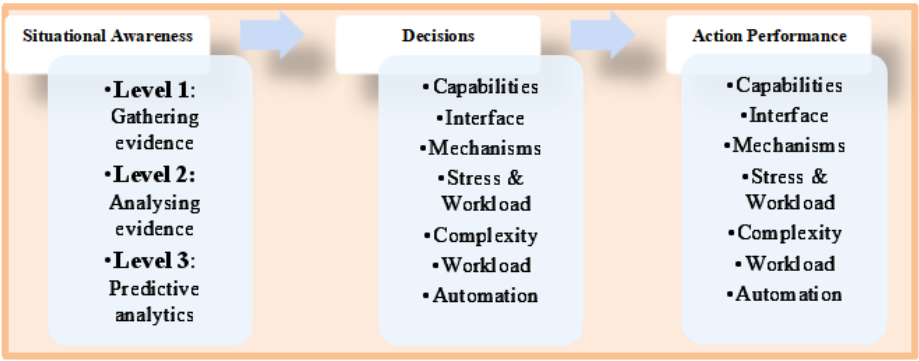
A biological neural network's properties are shared by an artificial neural network (ANN). The Feed Forward Neural Network (FFN), Convolutional Neural Network, and Recurrent Neural Network are members of the ANN family (RNN). Linear regression, Logistic regression, Linear Discriminant Analysis, Classification and Regression Trees, Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbor (K-NN), Kmeans clustering, Learning Vector Quantization (LVQ), Monte Carlo, Random Forest, Neural networks, and Q-learning are some of the more well-known examples of machine learning algorithms.

2.2.2 The future of AI in the fight against cybercrimes

New data architectures, analytical techniques, and tools are needed for big data analytics. To acquire risks from large data, analyze and filter information about these threats, and raise awareness of cybersecurity threats, we use the technique of "threat intelligence" (Sarker, I.H, et al, 2020). Figure 3 illustrates the situation awareness model, which consists of situation awareness, decisions, and action execution. Prior research has generally agreed that cybersecurity has developed into a

challenge for big data analytics. Additionally, even the data mining models that were previously utilized are no longer adequate to address the difficulties in cybersecurity (Hashem, I.A.T., 2015). The agility and robustness of a big data analytics model for cybersecurity can be used to judge it (Hashem, I.A.T., 2015).

Figure 3: Simplified Theoretical Model Based on Situation Awareness



2.2.3 Cybersecurity in Network Intrusion, Detection and Prevention System

With the aim of creating formalized models of data structures and data transmission from one object to another, info-communication is a natural science subject that analyzes the structure of objects and the process of interaction between these objects (Kuznetsov, N.A., 2005, p.1). The information technology and telecommunications sectors were initially impacted by the digital convergence process, which was adequately manifested in the unification of their technologies, the integration of their markets, and the harmonization of their regulatory frameworks (Sallai, G., 2012, p.2). As a result, different networks, services, and user terminals have been linked to different contents, and their marketplaces and regulatory frameworks have been governed separately. The term information and communications technology (ICT) is frequently used and typically describes the integration of the information and telecommunications technology sectors with the media technology sector based on standard digital technology.

Computer and communications security are two components of information systems security. Cybersecurity's total strength is

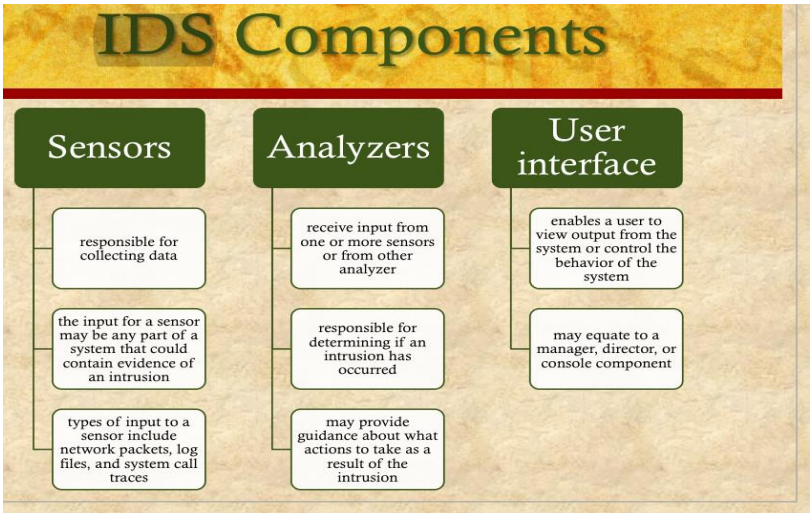
determined by its weakest link (Nielsen, R., 2015, p.8). The firm objectives must explicitly state access controls and security measures. It's crucial to restrict internet access for employees to those uses that are crucial to the business. However, the permissions granted must be carefully watched, particularly when accessing from outside the company's firewall. When it comes to network security, all traffic should be routed through a single point, and the firewall should only have the ports opened that are required for business activity. By offering VPN support, the network configuration can be reinforced (Nielsen, R., 2015, p.18). Network security calls for both an intrusion detection system (IDS) and an intrusion prevention system (IPS). However, with some legal guidance, organizational policies should outline the steps to handle information security.

The Intrusion Detection System (IDS) typically uses sensors, analyzers, and a user interface, as indicated on Figure 4 below. The following areas ought to be covered by the policies (Nielsen, R., 2015, p. 14):

- Personal Electronic Devices (PED)
- Acceptable Use
- Records Retention
- Identity Protection
- Server, Service and Project Computing Security
- Data Encryption

The IDS can either be network-based or host-based.

Figure 4: The IDS Components (Source: Stallings, W., 2015, p.6)



A firewall is essentially a computer server that communicates with external computer systems and protects important files on network PCs. It offers network protection against external threats (Stallings, W., 2015, p.10). Running Systems Operating systems are secured via hardening, which entails installing and patching the operating system before hardening and/or configuring the operating system to safeguard the system (Stallings, W., 2015, p.28). The following steps make up the creation and setup of the Bayesian network:

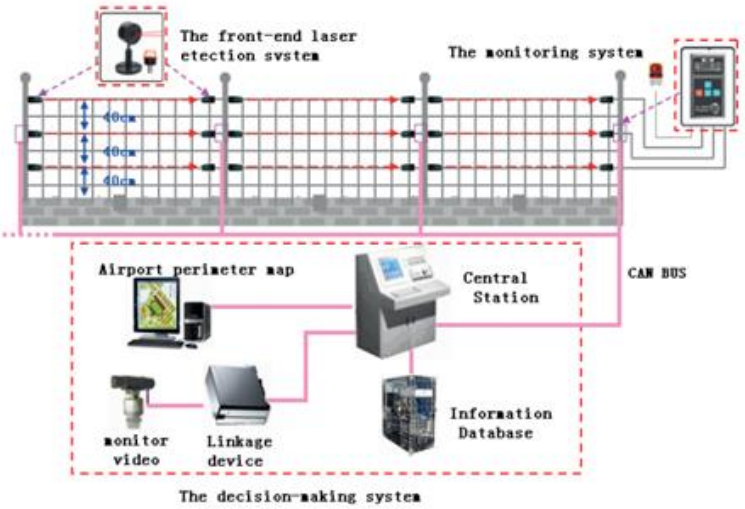
- Traffic sample obtaining to establish the information source in order to gather the sample
- Structural Learning, which defines the operational model
- Parametric Learning of the quantitative model
- Bayesian Inference
- Adaptation.

Longer-range coverage, a rise in bandwidth demand, and an increase in user density are anticipated benefits of next-generation access networks. Global adoption and deployment of optical fiber as the standard solution for next-generation fixed access networks. A probability model can effectively handle network intrusion and risk probability. To get around the issue with low-pass and high-pass filter-

based denoising techniques, the network signal is first processed using the principal component analysis (PCA) method (Wei and Liu, 2016). By projecting the initial variance of the data into the first and second coordinates, the orthogonalized linear transformation transforms the network signal into a new coordinate system (referred to as the first principal component and the second principal component, respectively). PCA can minimize the size of the network signal captured on the receiving device and remove background noise from the surroundings (Wu, 2018, p.2). By projecting the first variance of the data at the first and second coordinates, the network signal is translated into a new coordinate system using an orthogonalized linear transformation (referred to as the first principal component and the second principal component, respectively). PCA can minimize the size of the network signal gathered on the receiving device and remove noise from the surrounding environment (Wu, 2018, p.2).

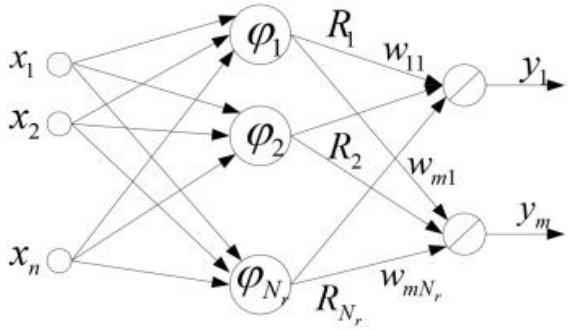
Infrared detection, vibration cables, subterranean cables, microwave detection, video surveillance, tension fencing, and other technologies are typically utilized for airport perimeter security, however laser detection technology is rarely employed (Wu et al, 2016, p.1). Wu et al. (2016) separated the front-end laser detection system into three parts: the decision-making system, the monitoring system, and the front-end laser detection system in account of the airport's surroundings, as shown on Figure 5.

Figure 5: The anti-intrusion laser alarm system (Source: Wu et al 2016, p.2).



At present, pattern recognition techniques are used to detect intrusion incidents at airport perimeters. These techniques include artificial intelligence, fuzzy theory, expert systems, neural networks, genetic algorithms, and other pertinent techniques. Artificial neural networks now have the unique capacity to interpret nonlinear data and store information in dispersed fashion (Wu et al, 2016, p.6). According to the network model in Figure 6, the radial basis function (RBF) neural network is built as a two-forward type neural network, and the value of the RBF is decided by the output of the intermediate layer node.

Figure 6: The structure of the Radial Basis Function (RBF) neural network (Source: Wu et al, 2016, p.6)



$X = \{x_1, x_2, \ldots, x_n\}$ are n-dimensional input vectors and the output of the hidden layer nodes are the RBF values. The input mapping to a new space is provided by the hidden layer unit which performs a nonlinear transformation. RBF is essentially a Gaussian function which is expressed as follows (Wu *et al*, 2016, p.6):

$$R_j = \varphi_j(X) = e^{-\|X - C_j\|^2/(2\sigma_{2j})}, \quad j = 1,2, \ldots, N_r$$

Systems for detecting network intrusions were developed to do so. The attacks or malicious activity can be identified by examining the network's packet contents. However, packet inspection is a difficult, resource-intensive procedure that is typically unachievable. The attacks are revealed by merging flow-based and graph-based processes, according to Karimpour et al. (2016, p. 1).

Table 4 below provides a comprehensive overview of these intrusion detection techniques based on the four categories mentioned above.

Table 4: Anomaly detection methods (Source: Karimpour et al (2016, p.3)

Method	Data type	Attack	Proposed system	Accuracy
Graph in time series	Flow-based	DDoS	Graph-based	94.2%
Dispersion graph	Flow-based	DDoS	Graph-based	100%
Using flow concept	Flow-based	Dictionary	Flow-based	99%
Graph clustering and local deviation coefficient	Packet-based	DoS, Scan	Graph-based	95.3%
Graph clustering and local deviation factor	Packet-based	DoS, Scan	Graph-based	97.2%
Packet heard analyzing	Packet-based	DoS, Scan	Packet-based	95.4%

By applying the flow and graph-clustering principles in a way that represented the nodes, the edges, and the weight of edges through the IPs, the flows, and the number of flows in the graph, respectively, Karimpour et al. (2016, p.3) were able to identify an assault in the network. The anomaly points can be identified by comparing the average weight of the clusters that the graph-clustering algorithm reaches over various time intervals and thresholds. The results of the study by Karimpour et al. (2016, p. 4) included 7 weeks of network traffic and 5 categories of assaults: DoS, scan, local access, user to root,

and data. These results are displayed on Table 5 below, which shows the quantity and types of attacks in each classed period.

Table 5. Various attack types and their descriptions (Source: Karimpour et al, 2016, p.4)

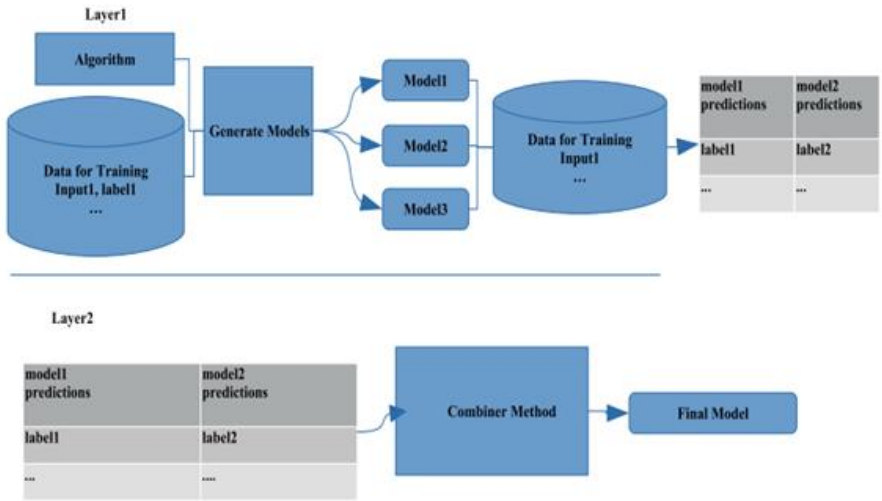
Attack type	Description
DoS	Denial of service; an attempt to make a network resource unavailable to its intended users: temporarily interrupt services of a host connected to the Internet
Scan	A process that sends client requests to a range of server port addresses on a host to find an active port
Local access	The attacker has an account on the system in question and can use that account to attempt unauthorized tasks
User to root	Attackers access a user account on the system and are able to exploit some vulnerability to gain root access to the system
Data	Attackers involve someone performing an action that they may be able to do on a given computer system, but that they are not allowed to do according to policy

The final model of attack detection can be created using the cluster-based data as input, and the proposed criterion is then calculated in the time series based on specified threshold points. As a result, the optimal threshold is found in the time series and derived from the detection rates using the advised method (Karimpour et al, 2016, p.5). For a maritime dispersed network, Li (2018) created a cooperative intrusion detection system using data mining techniques. an effective The developed distributed network intrusion model for marine environments was highly effective, consumed less memory, and had a detection rate of more than 92%. The marine distribution network, when employed in a ship navigation system, offers assurances for a ship's regular sailing. Researchers may be able to identify network infiltration using ML without the use of a signature database. By employing several classification algorithms as a combiner method, Demir and Dalkilic (2017) improved the model generation and selection techniques. Model generation was carried out using randomly chosen feature-filled subsets of the dataset, and the accuracy levels were higher than those of pure machine learning methods. The study had the highest detection rate for user-to-root attacks in comparison to other studies.

The same classification or regression problem is tackled in machine learning (ML) using ensemble learning, whose techniques produce a collection of models that are subsequently integrated. It was

demonstrated by Demir and Dalkilic (2017) that weak learners could receive support to become strong learners. The three most popular ensemble approach kinds are boosting, bootstrap aggregating, and stacking (Demir, N., and Dalkilic, G., 2017, p.1). Bootstrap aggregating is the process of training each model using samples taken at random from the training set. The random forest algorithm employs bagging and mixes random decision trees. By leveraging the misclassified training cases from the training of each subsequent model, boosting incrementally creates an ensemble model. Stacking, also referred to as stacked generalization, is a technique that makes use of an algorithm. Stacking is the generalization of existing ensemble methods and involves combining the outputs of forecasts from different models using an algorithm. Models are produced utilizing data and the training procedure. In the "model generation" phase, an algorithm with randomly chosen sub-datasets is used to build n models during the stacking implementation. As indicated on Figure 7 below, a two-layered training phase consists of a training set with each method and then the predicted labels of each model.

Figure 7: Training phase of tracking approach (Source: Demir, N., and Dalkilic, G., 2017, p.4)



A threat model created by Demir and Dalkilic (2017) gathers data at the packet level. In order to prevent assaults from being detected by

IDPSs or to get access to sensitive information on IDPSs such host configuration and known vulnerabilities, the IDPS components must first and foremost be secure.

3.0 RESEARCH METHODOLOGY

3.1 Overview

The research onion depicted in Figure 8 below served as a guide for the study philosophy, methodology, and design. This study used the Pragmatism paradigm, which is closely related to the Mixed Methods Research (MMR). Contextualization is necessary in order to address issues within the vast field of knowledge generation for strategic investment in STI with prospects for machine learning and cybersecurity.

The research methodology describes the procedures used to conduct the study and the justification for them. It is a method for thoroughly and methodically solving a research topic (Kotari, C.R., 2004). The pragmatic worldview served as the basis for the Mixed Methods Research technique. For the knowledge production stage, the researcher used a focus group discussion as the primary qualitative strategy. This was followed by a quantitative approach utilizing an experimental research design that entailed the creation of a Bayesian Network Model for Cybersecurity using the Snort platform. Because it was necessary to carefully characterize the details and traits of big data analytics models for cybersecurity, the researcher chose a descriptive study design. The purpose of the study was essentially an in-depth description of the models (Burt, D., et al, 2013). The researcher adopted a postmodern philosophy to guide the research. The study pointed out that there are regional and international differences in the definition, scope, and assessment of cybersecurity (Wilson, B.M.R., et al, 2015). Case studies have frequently been used in previous studies on cybersecurity (Wilson, B.M.R., et al, 2015).

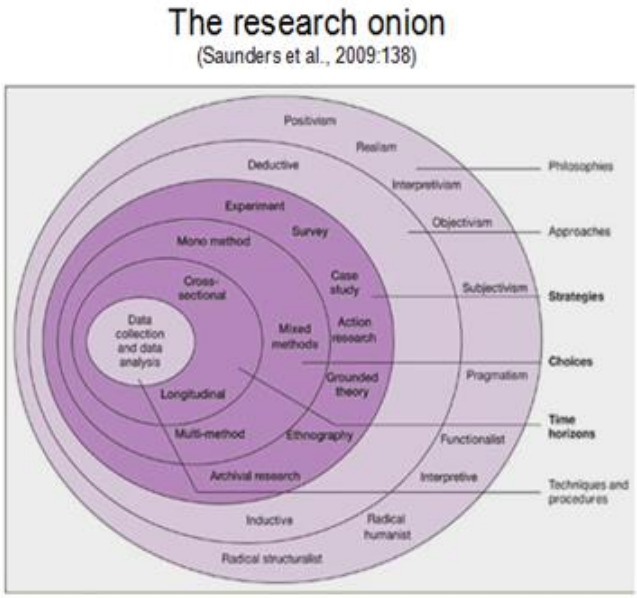


Figure 8. Research onion

Focused group discussions serve as the primary study design in the knowledge generation research program's predominantly qualitative research approach, which is based on an integrative research architecture that blends descriptive, narrative, theoretical, and experimental survey methodologies. Figure 9 below shows the integral research architecture, which combines descriptive methods, experimental and survey methods, theorizing methods, and narrative approaches. These techniques each correspond to one of the four (4) human modes: being, doing, knowing, and becoming. Empirical phenomenology (descriptive methods), storytelling (narrative methods), grounded theory (methodologies of theorizing), and case studies are the main techniques employed in integral research methods (experimental and survey methods).

The Integral Worlds approach emphasizes a comprehensive approach to research and social innovation that is based on the dynamics of the four worlds, South, East, North, and West, which are symbolic. Understanding the advantages and disadvantages of each world as well as the lessons that can be drawn from one another is beneficial. Each of the four worlds represents a certain region of the real world more

specifically in both metaphor and fact. As a result, the south has more direct ties to Africa, the east has ties to Asia, the north has ties to Europe, and the west has ties to America. The four realms are actually figuratively present in every society, every institution, and in every individual. A southern relational spirit of nature and community (human security), an eastern holistic spirit of culture and spirituality, a northern spirit of reason, and, ultimately, a western spirit of enterprise, structure, and continuity can all be found in every society, organization, and individual. Typically, one of the four worlds predominates in a given society, group, or person.

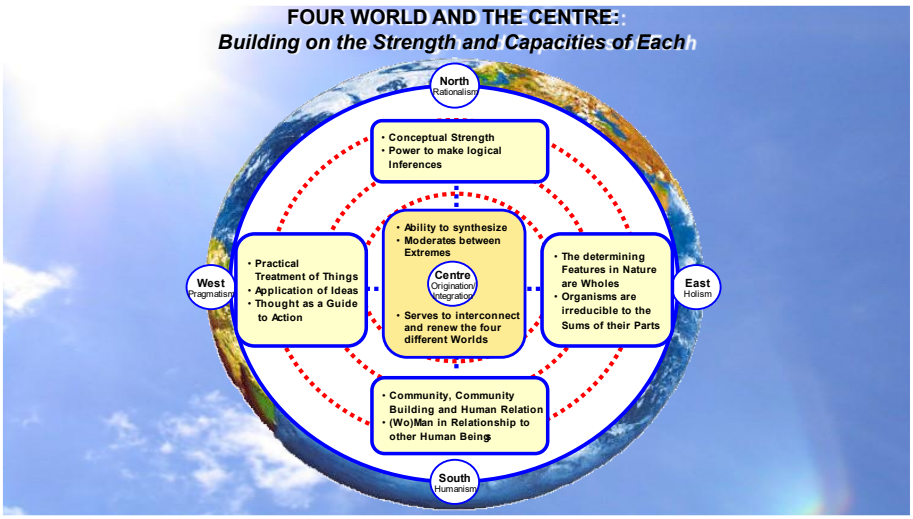


Figure 9: Integral Research Architecture

Focus Group talks served as the research design. The 22 institutions of knowledge generation in Zimbabwe that took part in the National Indaba on the National Science, Technology, and Innovation System of Zimbabwe, held on October 4, 2017, at the HICC in Harare, formed a Focus Group. This Working Group on Knowledge Generation had focus group meetings every week from October 4 through the end of November 2017.

The Work Plan for Each Milestone are Listed in Table 6 of the Work Plan for the Working Group on Knowledge Generation (the Focus Group) Using this Research Program.

Table 6: Work Plan for Knowledge Generation

Milestone	**Action step**
1 Knowledge Prioritisation	a) Identification of priority areas b) Engagement of institutions c) Identification of two technologies d) Map way forward for commercialisation
2 Advocacy of the National Science Technology and Innovation System of Zimbabwe	a) Develop Communication and publicity Framework b) Interfacing with the media for publicity
3 Knowledge co-creation and collaboration	a) Institution to institution collaboration b) Interministerial collaboration c) Institution to company collaboration d) Country to Country collaboration
4 Knowledge Acquisition	a) Diaspora engagement b) Formulate skills exchange programme in identified priorities c) Identify expired patents for utilisation
5 Knowledge Enterprenuerising and Commercialisation	a) Develop and adopt solid ideation protype process (SIPP) in 10 days b) Establish and maintain a national support (NSN)network in 10 days c) Develop and disseminate a national scheme to promote, incentives and recognise innovation d) Set up and maintain database of ongoing project works
6 Fusion and Adaption	a) Literature review b) Identification of technologies that will address priority areas c) Identify relevant local prayers who can take up the technologies d) Creation of database for commercially viable science e) End user experimentation of identified technologies for adaption f) Development of technology usage tracking mechanism

3.2 Quantitative data collection of the KDD'99 data set for the development a Bayesian Network Model

The research developed an effective network intrusion detection system using the KDDCup 1999 intrusion detection benchmark dataset. About 10 million records with 42 variables made up the population, which was the primary data retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (attributes). The information was retrieved from an archived source located at the University of California, Irvine, Irvine, CA 92697-3425, KDD Archive, Information and Computer Science. A sample of 494,020 records with 42 incidents were chosen from this population for data analysis.

3.3 Population, sampling and Model for analysis

A predetermined sample size of responders provides the researcher with the necessary data (Kumar, R., 2011). The population is the entire group of cases from which a sample was drawn (Saunders, et al, 2009). The term "population" refers to the entire group that the researcher desires to examine. The whole list of population members from whom a sample is obtained is referred to as the sampling frame, according to Saunders et al. (2009). Given that the population size is finite, the researcher used the Yamane's formula to calculate the sample size with a 95% confidence level (Saunders et al., 2009).

The purposive sampling method was applied to the 494,020 intrusion detection records from the KDD'99 Dataset, which served as the sample. Each item has a nonzero chance of being chosen from the population with an equal probability in probability sampling. Nonprobability sampling does not ensure equal inclusion of all participants or population units. The nonprobability sampling method would be most appropriate when the researcher encounters issues with limited resources, time, and labor. It can also be employed when the research does not seek to produce results that will be used to draw generalizations about the entire population. The purposeful selection of a participant based on their personal characteristics is known as the purposive sampling technique, also known as judgment sampling (Etikan, I., 2016, p.2).

The researcher determines what information is necessary based on expertise or experience, then goes out looking for persons who can and are willing to offer it. Purposive sampling, according to Etikan (2016, p. 2), is frequently employed in qualitative research to discover and choose the information-rich examples for the most effective use of the resources at hand. It frequently entails choosing certain groups or people as data sources. Purposive sampling of the Maximum Variation Sampling (MVS) kind was employed.

All data analytics models for cybersecurity that have been suggested and developed in works of literature, journals, conference proceedings, and working papers make up the research population for this study. From a survey of the literature and an analysis of the 8-person sample, the researcher discovered two data analytics models or frameworks. Interviews with a total of eight people were conducted. Even though there may not be much data, it will be enough for the purposes of this study at this time. In order to explore the usage of data analytics models in cybersecurity, the researcher analyzed secondary data. The researcher refers to the qualities of a perfect data analytics model for cybersecurity when examining the various data analytics models for cybersecurity. Big data, analytics, and insights make up the three main building blocks of the fundamental big data analytics paradigm for cybersecurity (Hashem, I.A.T., et al, 2015). The following Figure 10 shows this. The availability of big data on cybersecurity is the initial element of the bigdata analytics framework for cybersecurity. System logs and vulnerability scans are conventional sources of big data (Hashem, I.A.T., et al, 2015). Nevertheless, the sources of big data about cybersecurity have expanded to include computer- and mobile-based data, user-level physical data, human resources data, credentials, one-time passwords, digital certificates, biometrics, and social media data (Truong, T.C., 2020). Business mail, access control systems, CRM systems, human resources systems, a number of pullers in linked data networks, intranet/internet, industrial internet of things (IIoT), and IoT, collectors and aggregators in social media networks, and external news tapes are some of the basic sources of big data identified for cybersecurity work (Stallings, W., 2015). More reliable big data analytics models for cybersecurity have been developed using data mining and machine learning approaches to address the issues of big

data concerning cybersecurity (Hashem, I.A.T., et al, 2015). Big data analytics in cybersecurity uses support vector machine learning techniques, intrusion and malware detection methods, and data mining processes and algorithms (Hashem, I.A.T., et al, 2015). However, data nonstationarity, unbounded patterns, uniqueness, unequal time lags, high false alarm rates, and collusion attacks pose the biggest problems for intrusion detection systems (Menzes, F.S.D., et al, 2016). Big data analytics for cybersecurity must take a multi-layered and multidimensional strategy as a result. In other words, a big data analytics model for cybersecurity that is effective must be able to identify malware and intrusions at every level of the security architecture.

All data analytics models for cybersecurity that have been suggested and developed in works of literature, journals, conference proceedings, and working papers make up the research population for this study. From a survey of the literature and an analysis of the 8-person sample, the researcher discovered two data analytics models or frameworks. Interviews with a total of eight people were conducted. Even though there may not be much data, it will be enough for the purposes of this study at this time. In order to explore the usage of data analytics models in cybersecurity, the researcher analyzed secondary data. The researcher refers to the qualities of a perfect data analytics model for cybersecurity when examining the various data analytics models for cybersecurity. Big data, analytics, and insights make up the three main building blocks of the fundamental big data analytics paradigm for cybersecurity (Hashem, I.A.T., et al, 2015). The following Figure 10 shows this.

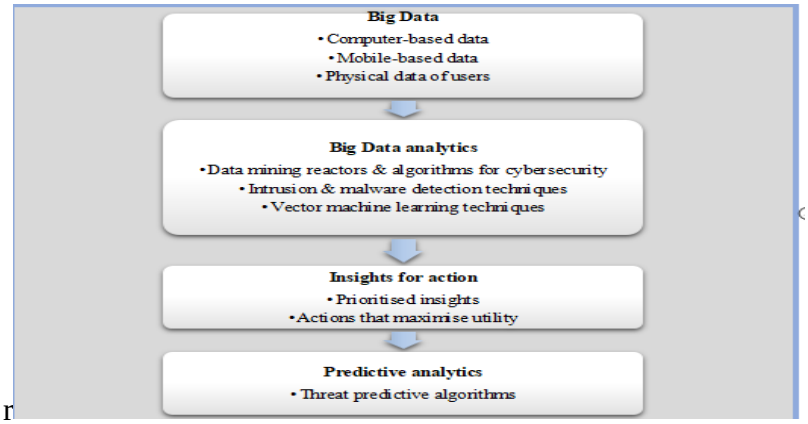


Figure 10: Big Data Analytics Model for Cybersecurity

The availability of big data regarding cybersecurity is the initial element of the big data analytics framework for cybersecurity. System logs and vulnerability scans are conventional sources of big data (Hashem, I.A.T., et al, 2015). Nevertheless, the sources of big data about cybersecurity have expanded to include computer- and mobile-based data, user-level physical data, human resources data, credentials, one-time passwords, digital certificates, biometrics, and social media data (Truong, T.C., 2020). Business mail, access control systems, CRM systems, human resources systems, a number of pullers in linked data networks, intranet/internet and industrial internet of things (IIoT)/IoT, collectors and aggregators in social media networks, and external news tapes are sources of big data about cybersecurity that have been identified by other researchers (Stallings, W., 2015).

More reliable big data analytics models for cybersecurity have been developed using data mining and machine learning approaches to address the issues of big data concerning cybersecurity (Hashem, I.A.T., et al, 2015). Big data analytics use malware and intrusion detection methods, data mining reactors and algorithms, and cybersecurity vector machine learning approaches (Hashem, I.A.T., et al, 2015). Data nonstationarity, collusion assaults, unbounded patterns, unequal time lags, individuality, and high false alarm rates are just a few of the problems intrusion detection systems must contend with (Menzes, F.S.D., et al, 2016). Big data analytics for cybersecurity must take a multi-layered and multi-dimensional approach as a result. In

other words, a big data analytics model for cybersecurity that is effective must be able to identify malware and intrusions at every level of the security architecture.

4.0 RESULTS AND ANALYSIS

4.1 Knowledge Generation for STI Investment Projects

Research produces information and technology, and innovation takes a step further by putting that knowledge to use in real-world applications. In order to create new ways of knowing, disciplinary traditions' value-adding material and methods are combined in a process known as knowledge co-creation. Applied research, idea/proof of concept development through commercialization, venture capital, and foreign direct investment are some examples of the creative support mechanisms (FDI). The possible contributing factors to why African scientific contribution globally is only 2% may include:

- Weakness or nonexistence of an environment advantageous for research;
- Deficient budget dedicated to research;
- Not rewarding status of the researchers;
- Rough evaluation of the impact of research on development.

The strategies for knowledge generation through the Rapid Results Initiative (RRI) are:

1. Advocacy of the National Science, Technology and Innovation System
2. Knowledge co-creation and collaboration
3. Knowledge acquisition
4. Knowledge prioritization
5. Knowledge fusion and adaption
6. Knowledge enterprenuerising and commercialization

The following important initiatives were recognized as important projects to be pursued in the upcoming 100 days after a study of the national research priorities, SDGs, and STISA 2024:

1. Post-harvest technologies
2. Small scale mining/mineral value addition/bio mining

3. Clean water alternatives
4. Tiles technologies from mining waste
5. ICT innovations for applications /Cyber security systems
6. Defence technologies (double use technologies, drones, puma vehicle, land mine detectors, etc.)

In order to determine whether the new innovations could be used in accordance with the major priority projects, a visit was conducted to Zimbabwe's higher education institutions. The most cost-effective technologies have been determined, and they are displayed on the schema below in Table 7:

Table 7: Key priority technological innovations in Zimbabwe

Priority Technology	**Schema**
1. Kwekwe Polytechnic Brick Machine	
2. Chinhoyi University of Technology Animal drawn LimeSpreader	

3. Harare Polytechnic Floor tile manufactured	

The National Development Strategy, which serves as the foundation for the Zimbabwean economy, aims to accomplish the following goals:

i. Improved revenue collection from important economic sectors, such as mining;
ii. Establishment of a Sovereign Wealth Fund;
iii. Improved revenue collection from important economic sectors, such as agriculture;
iv. Increased investment in infrastructure, including the development of energy and power, rail, roads, telecommunication, ICTs, aviation, water, and sanitation, through acceleration of the implementation of Public Private Partnerships (PPPs) and other private sector driven initiatives;
v. Increased Foreign Direct Investment (FDI);
vi. Establishment of Special Economic Zones;
vii. Continued use of the multi-currency system;
viii. Implementation of effective Value Addition policies and strategies; and
ix. Improved supply of electricity and water.

Therefore, the economic framework and scientific technological solutions deduced by Zimbabwe's national science, technology, and innovation system must serve as the foundation of the national development strategy. Special attention should be given to the following sectors of the economy:

1. Agriculture
2. Mining
3. ICT
4. Research and Development

5. Manufacturing
6. Construction
7. Human capital development
8. Health and social services
9. Defence and security

The financing model for the National Development Strategy has the following elements:

1) National resource mobilisation programmes from domestic resources driven by the Government of Zimbabwe
2) The establishment of the National Wealth and Innovation Fund
3) Collaboration with all the development partners in STEM-related projects and programmes
4) Public Private Partnerships (PPPs) in the proposed Special Economic Zones.

To strengthen the management of fiscal policy and the stability of the financial system, a number of fiscal reform initiatives are anticipated. Through a variety of measures, progress will be checked against the registered re-engagement process with creditors and international financial institutions (IFIs).

4.2 Performance of Machine Learning Algorithms

The obvious deficiencies of traditional security techniques have been shown. Therefore, it is important to identify practical answers for a dynamic and adaptable network defense system. Data mining algorithms need to be improved and optimized for classification of intrusion attacks. A comparison of the data mining methods that can be applied to intrusion detection is shown in Table 8.

Table 8: Advantages and disadvantages of data mining techniques (Source: Berman, D.S., et a, 2019)

Technique	**Advantages**	**Disadvantages**
Genetic Algorithm	❖ Finding a solution for any optimization problem ❖ Handling multiple solution search spaces	❖ Complexity to propose a problem space ❖ Complexity to select the optimal parameters ❖ The need to have local searching technique for effective functioning
Artificial Neural Network	❖ Adapts its structure during training without the need to program it	❖ Not accurate results with test data as with training data
Naive Bayes Classifier	❖ Very simple structure ❖ Easy to update	❖ Not effective when there are high dependency between features
Decision Tree	❖ Easy to understand ❖ Easy to implement	❖ Works effectively only with attributes having discrete values
K Mean	❖ Very easy to understand ❖ Very simple to implement in solving clustering problems	❖ Number of clusters is not automatically calculated ❖ High dependency on initial centroids.

In order to monitor computer systems and networks and establish whether an intrusion has occurred, intrusion detection systems (IDS) raise alerts as needed (Bloice, M., and Holzinger, A., 2018). Bloice, M., and Holzinger, A. (2018), on the other hand, addressed the issues with the Anomaly Based Signature (ABS), which lessens false positives by letting a user interact with the detection engine and raising classified alerts. The table, Table 3, below summarizes the benefits and drawbacks of ABSs and SBSs.

Table 3: Advantages and disadvantages of ABSs and SBSs models (Source: Bloice, M., and Holzinger, A., 2018).

Detection model	Advantages	Disadvantages
Signature-based	❖ Low false positive rate ❖ Does not require training ❖ Classified alerts	❖ Cannot detect new attacks ❖ Requires continous updates ❖ Training could be a thorny task
Anomaly-based	❖ Can detect new attacks ❖ Self-learning	❖ Propne to raise false positives ❖ Black-box approach ❖ Unclassified alerts ❖ Require initial training

The performance of each of the classical Machine Learning algorithms is presented below from Figure 11.

4.2.1 Classification and Regression Trees (CART)

The performance results of our CART algorithm in forecasting bank failure on the training set are displayed in Table 4 below. On the training dataset, the algorithm's accuracy rate was 82.8%. Our ideal model's best tuning or complexity parameter was 0.068. The method had a kappa of 88.72% and an accuracy level of 92.5% on the test dataset. Only 2 instances were incorrectly categorized as moderate and 1 as satisfactory by the algorithm.

TABLE 4: CART model performance.

Complexity Parameter	Accuracy	Kappa	AccuracySD	KappaSD
0.06849315	0.8275092	0.7519499	0.04976459	0.07072572
0.15753425	0.7783150	0.6683229	0.07720896	0.14039942
0.42465753	0.5222344	0.1148591	0.08183351	0.18732422

The accuracy of the CART model based on the complexity parameters of different test runs is shown on Figure 11 below. The complexity

parameter or the best tune parameter of 0.068 optimized the model performance.

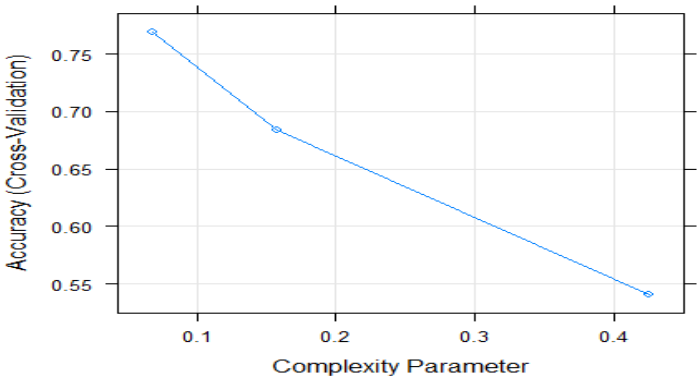


FIGURE 11: CART accuracy curve.

4.2.2 Support Vector Machine

According to Table 5, the SVM model's accuracy rate for forecasting bank solvency on the training dataset was 79.1%. Our extremely effective model's optimal tuning sigma and cost values were 0.05 and 1, as shown in Figure 12 below. The Kappa statistic and the Kappa SD were, respectively, 67.9% and 0.13. The method had an accuracy level of 92.5% and a kappa of 88.54% on the test dataset. Comparing the method to the CART algorithm, the algorithm only misclassified 3 instances as moderate.

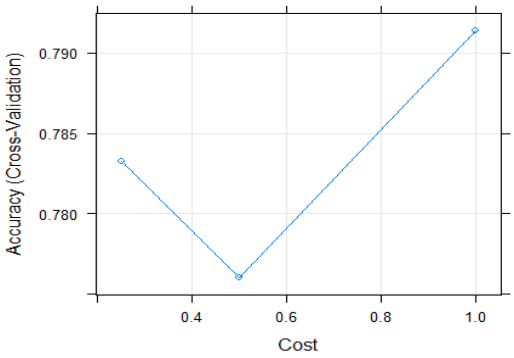


FIGURE 12: SVM accuracy curve

TABLE 5: Support Vector Machine performance

sigma	c	Accuracy	Kappa	AccuracySD	KappaSD
0.050398	0.25	0.783223	0.678536	0.095598	0.140312
0.050398	0.50	0.776007	0.661354	0.087866	0.132552
0.050398	1.00	0.791391	0.678694	0.080339	0.126466

4.2.3 Linear Discriminant Algorithm

TABLE 6: Linear Discriminant algorithm performance

Accuracy	Kappa	AccuracySD	KappaSD
0.8042399	0.7038131	0.1016816	0.159307

As shown in table 6, the LDA attained an accuracy level of 80% on the training dataset. The Kappa SD was 0.16 and the Kappa statistic was 70%. The algorithm's accuracy level and kappa on the test dataset were both 90%. Only 4 instances that performed poorly compared to the CART method were incorrectly labeled as moderate by the algorithm.

4.2.4 K-Nearest Neighbor

Table 7 shows the K-NN algorithm performance and confusion accuracy on Figure 10.

K	Accuracy	Kappa	AccuracySD	KappaSD
5	0.5988645	0.3698931	0.1280376	0.2158109
7	0.6268864	0.4072928	0.1564920	0.2703504
9	0.6621978	0.4715556	0.1747903	0.2881390

TABLE 7: K-NN algorithm performance

The training dataset's accuracy rate was 66.2%. According to the accuracy curve in Figure 13 below, k=9 or 9 neighbors was the ideal tuning parameter for our model. The Kappa statistic was 47.2%, and the Kappa SD was 0.17. The algorithm's accuracy level and kappa on the test dataset were 67.5% and 49%, respectively. Compared to other methods, the algorithm was not very good in classifying bank performance.

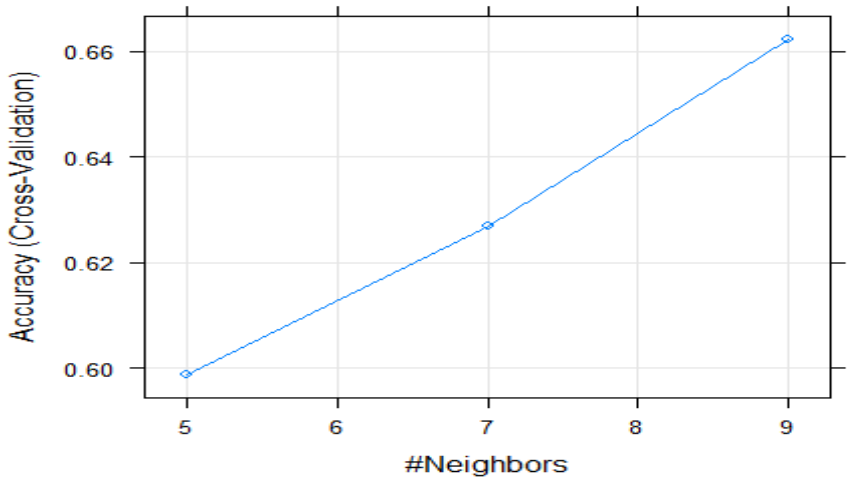


FIGURE 13: K-NN confusion accuracy graph

4.2.5 Random Forest

mtry	Accuracy	Kappa	AccuracySD	KappaSD
2	0.8272527	0.7421420	0.10396454	0.15420079
14	0.8554212	0.7829891	0.06069716	0.09303130
16	0.8482784	0.7718935	0.06455248	0.09881991

TABLE 8: Random Forest performance

Our random forest's accuracy on the training set was 85.5%, as shown in table 8. The number of predictors chosen at random for the purpose of building trees, as illustrated in Figure 14, that was selected as the optimal tuning parameter for our model was 14. The Kappa statistic was 78.3%, and the Kappa SD was 0.09, respectively. The algorithm's accuracy level and kappa were both 96% on the test dataset. Compared to other algorithms, the algorithm was quite good in classifying bank performance.

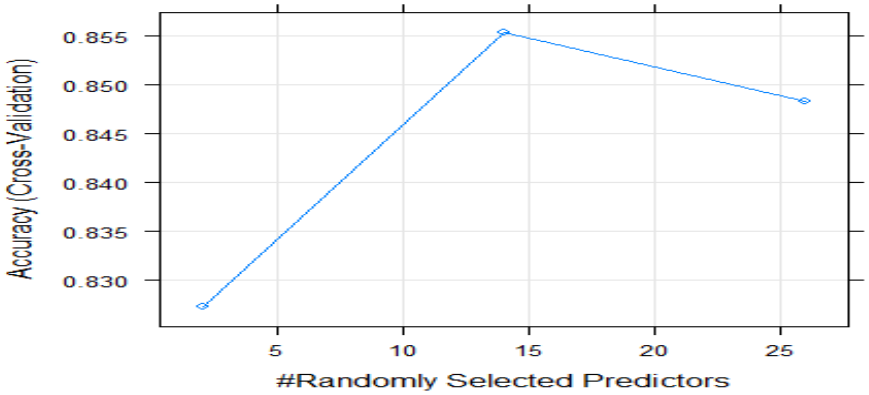


FIGURE 14: Random Forest accuracy graph

4.2.6 Challenges and Future Direction

Additionally, it means that data submission to the Reserve Bank of Zimbabwe continues to grow exponentially as the number of financial activity rises. This difficult situation offers limitless chances to use neural network-based deep learning (DL) methodologies to anticipate the solvency of Zimbabwean banks, especially in light of recent developments in machine learning (ML) and artificial intelligence (AI). In order to build a strong early warning supervisory tool based on big data analytics, machine learning, and artificial intelligence, future work will concentrate on finding more variables that could potentially result in subpar bank performance and incorporating these into our models.

The researcher analyses the two models that have been proposed in literature with reference to an ideal data analytics model for cybersecurity presented in Section 3.

4.2.7 Model 1: Experimental/ Prototype Model

In the first instance, the researcher takes use of the Stallings, W. (2015) model, which, despite being created in the context of the public sector, may be extended to businesses in the private sector. The key features of the experimental model are outlined in Table 9 below. Readers are directed to the prototype model that is also shown in Stallings, W. (2015).

Table 9: Experimental Big Data Analytics Model for Cybersecurity

MODEL ATTRIBUTES	DESCRIPTION
HBase working on HDFS (Hadoop Distributed File System)	• HBase, a non-relational database, facilitates analytical and predictive operations • Enables users to assess cyber-threats and the dependability of critical infrastructure
Analytical data processing **module**	• Processes large amounts of data, interacts with standard configurations servers and is implemented at C language • Special interactive tools (based on JavaScript/ CSS/ DHTML) and libraries (for example jQuery) developed to work with content of the proper provision of cybersecurity
Special interactive tools and libraries	• Interactive tools based on JavaScript/ CSS/ DHTML • Libraries for example jQuery developed to work with content for • Designed to ensure the proper provision of cybersecurity
Data store for example (MySQL)	• Percona Server with the ExtraDB engine • DB servers are integrated into a multi-master cluster using the Galera Cluster.
Task queues and data caching	• Redis
Database servers balancer	• Haproxy
Web server	• nginx , involved PHP-FPM with APC enabled
HTTP requests balancer	• DNS (Multiple A-records)
Development of special client applications running Apple iOS	• Programming languages are used: Objective C, C++, Apple iOS SDK based on Cocoa Touch, CoreData, and UIKit.
Development of applications running Android OS	• Google SDK

Software development for the web platform	• PHP and JavaScript.
Speed of the service and protection from DoS attacks	• CloudFare (through the use of CDN)

(Source: Stallings, W., 2015).

4.2.8 Model 2: Cloud computing/Outsourcing

In the second approach, a company hires a cloud computing service provider to manage its data. Because they specialize in data and networks, cloud computing service providers typically have superior big data analytics models, advanced detection and prediction algorithms, better state-of-the-art cybersecurity technology, and better protocols. It should be underlined, nonetheless, that providers of cloud computing services are not immune nor exempt from cyber-threats and attacks.

4.3 Development of the Bayesian Network Model

Planning, generalization, and prediction are all possible with Bayesian networks. It must be emphasized that depending on the type of service offered from each unique piece of equipment, i.e., from each different IP address and from each different TCP destination port, network traffic behavior as well as payload protocol lexical and syntactical patterns may fluctuate significantly. The usage of a multi-instance structure with several Dynamic Bayesian Networks, one for each combination of TCP destination address and port, was suggested by Bringas, P.B., and Santos, I. (2010). It must be able to concurrently give an effective reaction against both well-known and zero-day assaults, according to Bringas, P.B., and Santos, I. (2010). Bayesian networks demand a lot of processing power. In order to speed up the process, many of the jobs that need to be completed must be designed in a parallel manner (Bringas, P.B., and Santos, I., 2010, p.240).

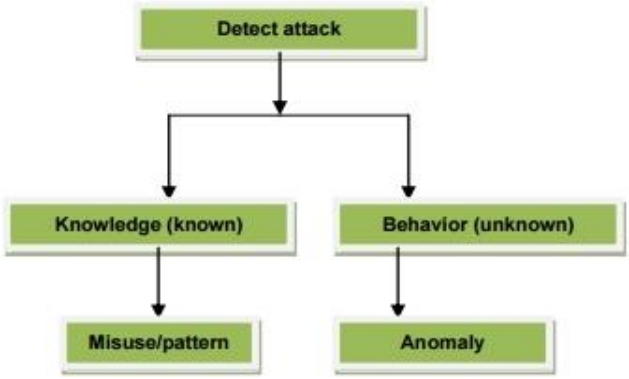
The Network Intrusion Detection System's overall behavior can be modified to accommodate unique requirements or configurations, although doing so requires a high level of complexity in the Bayesian structures and conditional probability parameters. To prevent denial of

service attacks, the dynamic regulation of knowledge representation models can be carried out by using sensibility analysis, automatically enabling or disabling expert modules using a single combined heuristic measure that takes into account specific throughputs and representative features (Bringas, P.B., and Santos, I., 2010, p.242). Additionally, it is possible to carry out model optimization in order to get the smallest number of representative parameters and smallest set of edges between them, which will improve overall performance. Approximate evidence propagation techniques can also be used to enhance inference and response time adaption.

A firewall and antivirus both function differently than an intrusion detection system (IDS). Firewall and antivirus software are bypassable and ineffective at preventing both internal and external threats. Firewalls often filter traffic using static rules rather than having the capacity to detect intrusions. IDS recognizes intrusion after its initial occurrence to stop similar assaults in the future (Murugan, S., and Rajan, M.S., 2014, p.1). The straightforward guidelines for attack analysis are depicted in Figure 15 below. The anomaly-based intrusion detection method's IDS sounds an alarm whenever there is any odd behavior on the network.

Figure 15: Analysis of Attack (Source: Murugan, S., and Rajan, M.S., 2014, p.2)

.

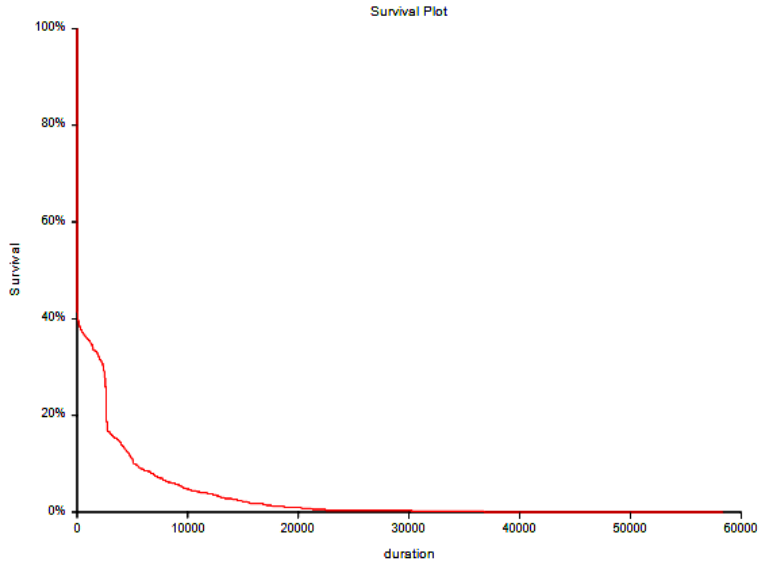


The secrecy and access control of the stored data are of great interest to defense security agencies and other militarily oriented companies. Therefore, it is crucial to research intrusion detection systems (IDS) to

identify and stop cybercrimes and safeguard these systems (Alocious, C., et al, 2014, p.1). In order to render a system inaccessible, distributed denial of service (DDoS) assaults bombard the server's network and end user systems with fictitious generated traffic. Legitimate users would not be able to access system resources in this way. By comparing a set of criteria designed to justify the user's conduct, signature-based detection, also known as rule-based detection, ascertains the user's behavior. An attack database with known signatures is what makes up a signature-based IDS. These assaults have been predefined based on attack analysis. Modern cyberattacks are expanding astronomically, and their organizational structures are constantly shifting. An evaluation and natural selection are the fundamental ideas of a genetic algorithm, which is a computational model. This means that in the course of natural selection, only the strongest will survive. In order to employ genetic algorithms, a set of rules for network data must be created. The application will read network packets in sniffer mode and show them on the console. The application will log packets to the disk while in packet logger mode. The study's findings were demonstrated using the SNORT IDS mode.

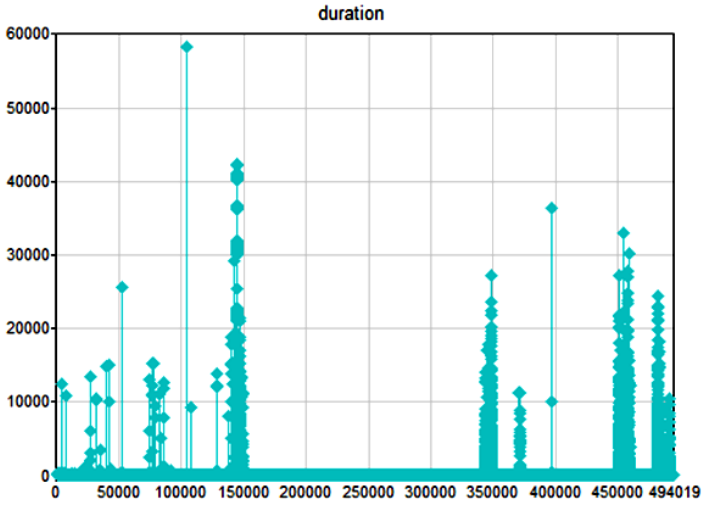
A non-parametric Survival Analysis was conducted with respect to one variable, ***duration***, and the result is shown as Kaplan-Meier Survival Curves on Figure 16 below.

Figure 16: Kaplan-Meier Survival Curve(s)



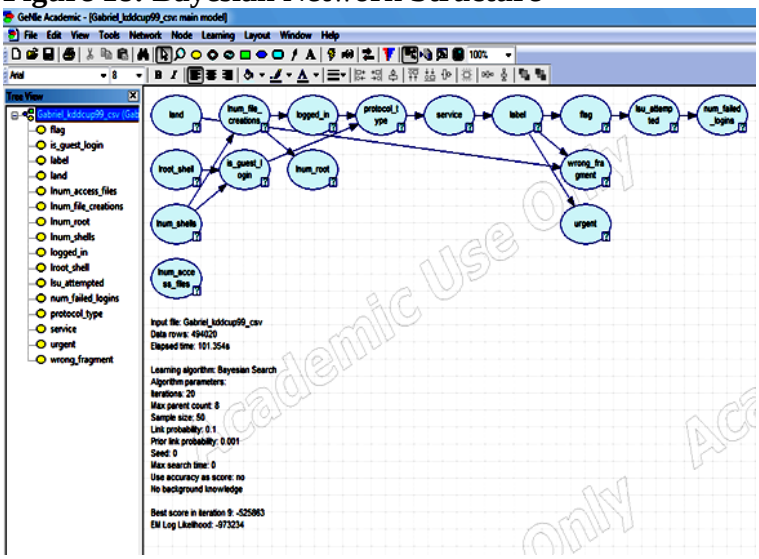
Time Series analysis of the variable ***Duration*** is shown on Figure 17 below.

Figure 17: Time Series of the variable Duration



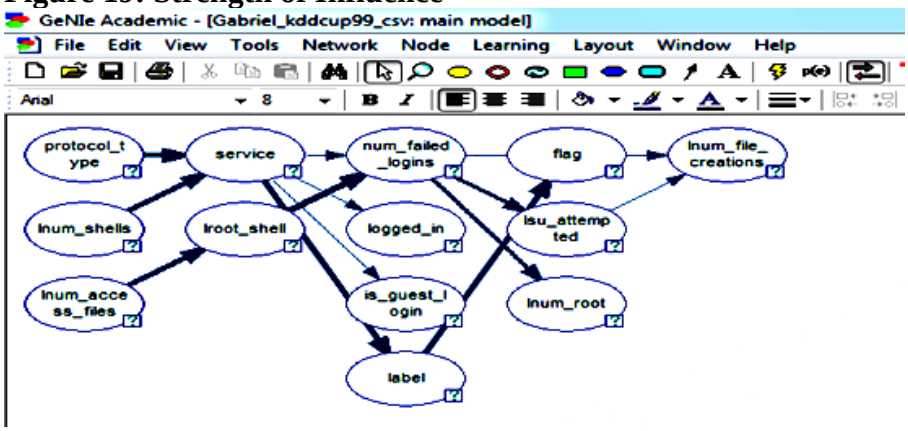
A new Bayesian Network model was created from the dataset and is shown on Figure 18 below.

Figure 18: Bayesian Network Structure



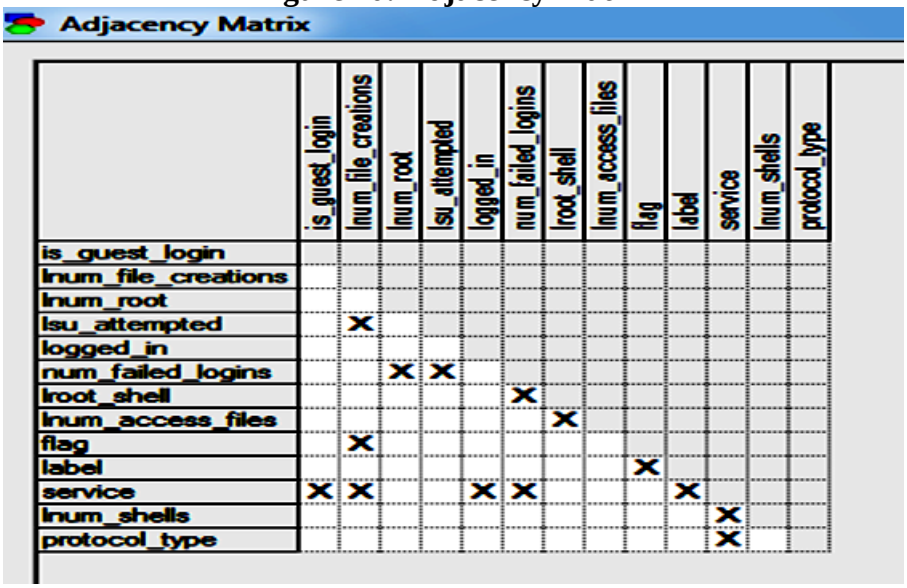
The Strength of Influence of the Bayesian Network is shown on Figure 19 below.

Figure 19: Strength of Influence



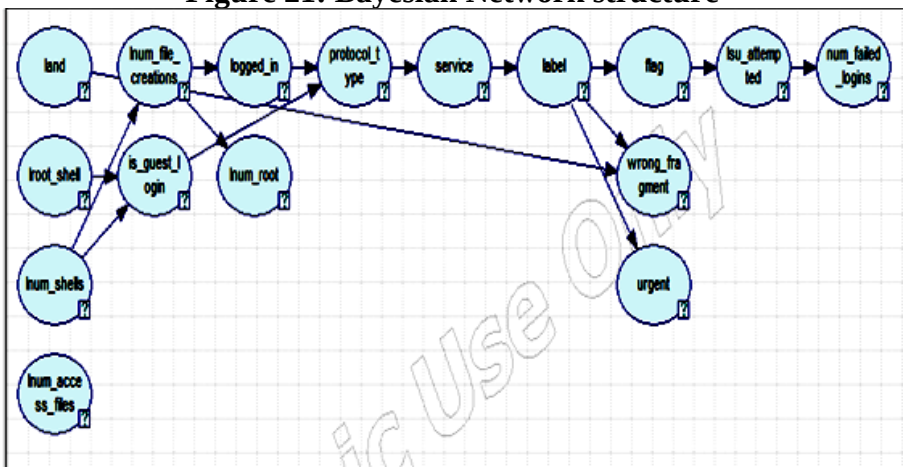
The Adjacency Matrix was computed for the network and is shown on Figure 20 below.

Figure 20: Adjacency Matrix



The Bayesian Network structure derived from the dataset is shown below on Figure 21.

Figure 21: Bayesian Network structure



The usage of Bayesian networks has a number of issues, one of which is the lack of correlation between the graphical structure and the corresponding probabilistic structure, which allows us to eliminate all the inference issues in graph theory. The operation for transforming the causal graph into a probabilistic representation has another issue. The Naive Bayes, a two-layer Bayesian network that presupposes total independency between the nodes, is one method in which BNs have been used in anomaly identification.

The mean values of the important variables are displayed on the table below in Table 10, which is based on a sample dataset of 494,020 occurrences with 42 variables analyzed.

Table 10: Mean values of the selected key variables

Variable	Mean
duration	47.9794
protocol_type	tcp
service	http
flag	SF
src_bytes	3025.62
dst_bytes	868.531
land	4.45E-05
wrong_fragment	0.00643294
urgent	1.42E-05
hot	0.0345188
num_failed_log+	0.000151816
logged_in	0.148245
lnum_compromis+	0.0102121
lroot_shell	0.000111332
lsu_attempted	3.64E-05
lnum_root	0.0113518
lnum_file_crea+	0.00108295
lnum_shells	0.000109307
lnum_access_fi+	0.00100806
is_guest_login	0.00138658
count	332.286
srv_count	292.907

serror_rate	0.176687
srv_serror_rate	0.176609
rerror_rate	0.0574335
srv_rerror_rate	0.0577191
same_srv_rate	0.791547
diff_srv_rate	0.0209824
srv_diff_host_+	0.0289962
dst_host_count	232.471
dst_host_srv_c+	188.666
dst_host_same_+	0.753781
dst_host_diff_+	0.0309058
dst_host_same_+	0.601936
dst_host_srv_d+	0.00668351
dst_host_serro+	0.176754
dst_host_srv_s+	0.176443
dst_host_rerro+	0.0581177
dst_host_srv_r+	0.0574118
label	normal

A site's introduction of new services or modifications to its policy are two examples of situations where anomalous behavior may occur. The advent of hybrid detection, which leverages misuse detection to have a high detection rate on known assaults and capacity to detect unexpected attacks, is the answer to these two issues. The most popular kind of hybrid system combines an anomaly detection and a misuse detection. It is possible that a hybrid IDS can be utilized by combining the misuse detection and anomaly detection modules, with the misuse detection module using the random forest method first to identify known intrusions. Evaluations with a portion of the KDDCUP'99 data set, which was employed in this study, revealed that the abuse detection module produced a high detection rate with a low false positive rate, while the anomaly detection component had the ability to discover novel intrusions.

A two-stage hybrid intrusion detection and visualization system that combines the benefits of signature-based and anomaly detection techniques was proposed by Zekrifa, D.M.S. (2014, p. 17), and it has

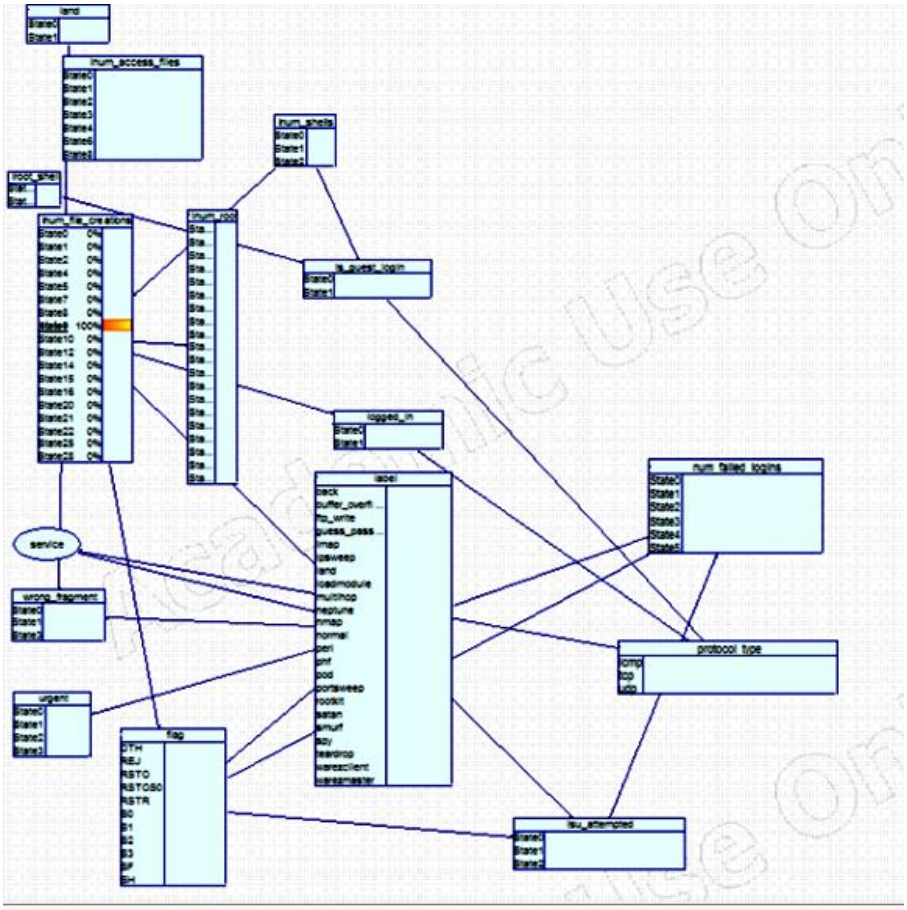
the capacity to detect both known and unidentified attacks on system calls. A novel hybrid IDS system made up of an anomaly detection module, a misuse detection module, and a decision support system is suggested as an enhanced IDS. The outcomes of the two earlier detection modules would be combined using the decision support system. As each anomaly detection technique has a varied detection capability, some additional hybrid systems combine several anomaly detection systems in accordance with certain criteria rather than mixing signature detection techniques and anomaly detection techniques. The hybrid system's goals include maintaining a respectable detection rate and lowering the high number of false warnings produced by the present anomaly detection methods (Zekrifa, D.M.S., 2014, p.19).

The cybersecurity challenges that are being faced in developing countries, like Zimbabwe, include the following:

1. Infrastructure
2. Legal frameworks
3. Harmonization of legislation
4. Balancing harmonization and country specific needs
5. Systems
6. Education and awareness
7. Cybersecurity knowledge
8. Affordability and funding
9. Perceived low susceptibility to attacks
10. Lack of adequate frameworks that speak to their cybersecurity needs
11. Reporting cybercrime
12. Data sharing

The Bayesian Network Model developed is shown on Figure 22 below.

Figure 22: The Bayesian Network Model developed



The vast majority of network security measures currently in use are unable to stop assaults on distributed computer systems because they are dynamic and getting more complicated. Building an automatic and adaptive defensive mechanism for computer networks becomes important as a result. Encryption and firewalls are the first methods currently used to prevent invasions. Next comes intrusion detection system (IDS) technology, which may identify illegal access to and abuse of computer systems by both internal users and external offenders (Tran, T.P., 2009, p.iv). Artificial Neural Networks (ANN) are an example of an artificial intelligence (AI) technology that has

been used to enhance detection performance. However, ANN requires expensive computation.

Learning the Bayesian network's structure and utilizing data and previous knowledge are necessary for updating the probability in the network structure (Soberanis, I.V.D., 2010, p.66). The learning mechanism in sequential updating of Bayesian Networks gets the data as a stream of observations, and it produces a model based on the data observed up to that point. There are several Sequential Update methods, including the incremental, maximum a posteriori probability (MAP), and naïve methods (Soberanis, I.V.D., 2010, p.67). The massive amount of data does, however, necessitate a lot of memory. The MAP approach assumes that the data being summarized has a probability distribution depending on the present model in order to cope with the problem of big data sets by storing all the prior data by summarizing the data utilized in the model up to this point. Both incremental and recursive Bayesian updating is possible. Recursive bayesian updating is amazing because it is straightforward and has so many uses. The junction tree technique offers a systematic and effective approach of clustering. With this approach, a junction tree, an updated graph, is subjected to bayesian propagation. Cycles in a network are removed by the Junction tree method by grouping them into single nodes (Soberanis, I.V.D., 2010, p.70). By updating the probabilities, which entails using fresh data or evidence to calculate the posterior probability distributions, the Bayesian network method of reasoning is used. The computation of the posterior probability distribution for a set of query nodes, given values for some evidence nodes, is known as Bayesian updating for any probabilistic inference.
Utilizing prior information, or what we may refer to as the training data, might aid in learning. Prior knowledge can really be very helpful when learning. The information we gather or are given can significantly speed up the decision-making process. Depending on the data, a variety of learning strategies may be used. The learning strategy can be reinforced, unsupervised, or both. The modification of the network's state in response to information supplied by the environment is known as supervised learning (Soberanis, I.V.D., 2010, p.74). In unsupervised training, the training data is given to the network and the likely or unlikely data is derived; however, the desired outputs are not provided.

The system must then choose the features it will employ to organize the input data, or the network must interpret the inputs on its own. In order to extract the necessary set of features, Soberanis, I.V.D. (2010, p. 126) suggested an online traffic categorization approach that uses the unigram payload distribution model. Following that, the J48 decision tree is used to categorize the network applications according to unigram features, and it is seen that the signatures are present in a few specific locations in the payload. Through a weighted scheme over the characteristics utilizing a genetic algorithm, it is crucial to give the features that appear in these more significant positions greater weight.

Two major issues were found by Almutairi, A. (2016). The first is that signature-based intrusion detection systems, like SNORT, are unable to automatically identify attacks with fresh signatures. The second issue is with multi-stage assault detection; it has been discovered that signature-based methods are ineffective in this regard. The first problem was solved by Almutairi, A. (2016), who created a multi-layer categorization system. A decision tree served as the foundation for the first layer, while a hybrid module that employs both fuzzy logic and neural networks as data mining approaches served as the source for the second layer. In the event that the first layer is ineffective at detecting new attacks, the second layer was created. The SNORT signature holder is automatically updated by this system after it recognizes attacks with fresh signatures. The outcomes demonstrated that attacks with fresh signatures had a high detection rate. The false positive rate needs to be reduced, it has been noted. The second problem was addressed by using fuzzy logic to the evaluation of IP data. Instead of focusing on the order and substance of the traffic, this technique examined the identities of the participants. The outcomes demonstrated that in some cases, this technique can assist in extremely early attack prediction. Almutairi, A. (2016) acknowledged that combining this strategy with a different approach—such as event-correlation—that examines the sequencing and contents of the traffic will result in greater performance than each approach used alone. The tremendous growth in the volume and complexity of the data that needs to be analyzed, however, is one of the biggest obstacles to developing an efficient solution using data mining. Due to the high cost of calculation caused by this, data mining may use a lot of CPU and memory

resources that are either expensive or unavailable. Therefore, conducting network traffic analysis utilizing a selection of the data rather than all of them for the aim of creating profiles may result in incorrect results.

5.0 CONCLUSION

It is envisaged that the national innovation programme would:

- make Zimbabwe's innovation system truly international, by supporting partnerships, collaboration and foreign investment in Zimbabwean R&D;
- build a culture of innovation and new ideas by strengthening investment in creativity and knowledge generation;
- accelerate the take up of new technology, so Zimbabwean firms can access the best ideas from around Zimbabwe and the rest of the world;
- focus incentives for business R&D to promote global competitiveness, delivering the best outcomes for exports and economic growth; and
- Enable resource mobilisation for the specific national innovations and industrialisation programmes which are STEM-related.

Through a variety of communication and visibility events, media coverage, social media profiles, promotional materials, and research publications and bulletins, advocacy and publicity work was accomplished.

As a subset of artificial intelligence, machine learning algorithms can be categorized as supervised, unsupervised, semi-supervised, and reinforcement learning algorithms. Large data sets are automatically analyzed for patterns and relationships, which leads to the creation of models for those patterns. Big data analytics focuses on the volume, diversity, and velocity of data in addition to the size of the data.

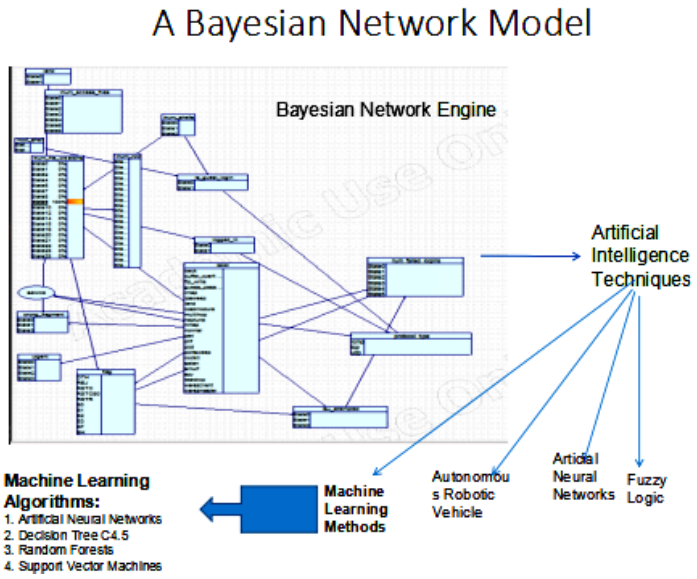
While a review of the literature revealed that different big data analytics models are used by institutions and nations for cybersecurity, the researcher also showed that, despite their varied requirements, these models have many important things in common. For instance, reactors

and detection algorithms are typically present in all models but vary in terms of complexity. Additionally, it is important to remember that many small businesses will typically follow Model 2, whereas very large businesses and delicate public sector institutions will typically adopt Model 1. Although the framework used to construct a data analytics model for cybersecurity in a cloud computing services provider may share similar characteristics with that produced by an institution on its own, this may also help to explain why the models employed may differ.

The researcher provided two approaches for using data analytics to cybersecurity in this section. In the first experimental or prototype model, an institution designs and implements a prototype, while in the second model, services offered by cloud computing businesses are used. Future research is expected to concentrate on ML applications and novel algorithmic performance as well as responsible AI.

The final Bayesian Network model developed is shown on the diagram below on Figure 23.

Figure 23: The Final Bayesian Network model



However, in order for network detection and prevention systems like autonomous robotic vehicles, artificial neural networks, and fuzzy logic to work, the Bayesian Network must be supported by artificial intelligence paradigms. Furthermore, these algorithms ought to be used in the basic network intrusion detection and prevention system:

- Support Vector Machines,
- Artificial Neural Network,
- K-Nearest Neighbour,
- Naive-Bayes and
- Decision Tree Algorithms

Use of machine learning methods, particularly Artificial Neural Networks (ANN), Decision Tree C4.5, Random Forests, and Support Vector Machines, provides alternative, improved solutions (SVM). However, using Bayesian networks has its own drawbacks, such as the inability to fully resolve all of the inference issues in graph theory due to a lack of correlation between the graphical structure and associated probabilistic structure. However, due to their complexity, these issues have generated a lot of research. The process of transforming the causal graph into a probabilistic representation is likewise difficult.

References

Berman, D.S., Buczak, A.L., Chavis, J.S., and Corbett, C.L. (2019). "Survey of Deep Learning Methods for Cyber Security", *Information* **2019**, *10*, 122; doi:10.3390/info10040122.

Bloice, M. & Holzinger, A., (2018), *A Tutorial on Machine Learning and Data Science Tools with Python.* Graz, Austria: s.n.

Burt, D., Nicholas, P., Sullivan, K., & Scoles, T. (2013). Cybersecurity Risk Paradox. *Microsoft SIR*.

Government of Zimbabwe's Various Budget Statements by the Ministry of Finance, accessed at http://www.zimtreasury.gov.zw/

Government of Zimbabwe's Various Monetary Policy Statements by the Reserve Bank of Zimbabwe, accessed at http://www.rbz.co.zw/

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. In *Information Systems*. https://doi.org/10.1016/j.is.2014.07.006.

Jones, C. (1998) *Introduction to Economic Growth,* (W.W. Norton, 1998 First Edition, 2002 Second Edition).

Kabanda, G. (2013), ***"****African context for technological futures for digital learning and the endogenous growth of a knowledge economy****",*** Basic Research Journal of Engineering Innovation (**BRJENG**), Volume 1(2), April 2013, pages 32-52, http://basicresearchjournals.org/engineering/pdf/Kabanda.pdf

Kabanda, G., (2021), *"Performance of Machine Learning and Big Data Analytics paradigms in Cybersecurity and Cloud Computing platforms*", Global Journal of Computer Science and Technology: G Interdisciplinary Volume 21, Issue 2, Version 1.0, Year 2021; Type: Double Blind Peer Reviewed

International Research Journal; Publisher: Global Journals Online ISSN: 0975-4172 & Print ISSN: 0975-4350; Performance of Machine Learning and Big Data Analytics Paradigms in Cybersecurity and Cloud Computing Platforms (globaljournals.org).

Kothari, C.R. (2004). *Research Methodology Methods and Techniques 2nd Revised Edition*. New Age International Publishers

Mazumdar, S., and Wang, J., (2018). *Big Data and Cyber security: A visual Analytics perspective in S. Parkinson et al (Eds),* Guide to Vulnerability Analysis for Computer Networks and Systems.

Menzes, F.S.D., Liska, G.R., Cirillo, M.A. and Vivanco, M.J.F. (2016) Data Classification with Binary Response through the Boosting Algorithm and Logistic Regression. Expert Systems with Applications, 69, 62-73. https://doi.org/10.1016/j.eswa.2016.08.014

Murugan, S., and Rajan, M.S., (2014). Detecting Anomaly IDS in Network using Bayesian Network, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727, Volume 16, Issue 1, Ver. III (Jan. 2014), PP 01-07, www.iosrjournals.org

Napanda, K., Shah, H., and Kurup, L., (2015). *Artificial Intelligence Techniques for Network Intrusion Detection, International Journal of Engineering Research & Technology (IJERT),* ISSN: 2278-0181, IJERTV4IS110283 www.ijert.org, Vol. 4 Issue 11, November-2015.

Nielsen, R. (2015). CS651 Computer Systems Security Foundations 3d Imagination Cyber Security Management Plan, Technical Report January 2015, Los Alamos National Laboratory, USA.

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from

machine learning perspective. *Journal of Big Data*. https://doi.org/10.1186/s40537-020-00318-5

Siti Nurul Mahfuzah, M., Sazilah, S., & Norasiken, B. (2017). An Analysis of Gamification Elements in Online Learning To Enhance Learning Engagement. *6th International Conference on Computing & Informatics*.

Stallings, W., (2015). Operating System Stability. Accessed on 27th March, 2019. https://www.unf.edu/public/cop4610/ree/Notes/PPT/PPT8E/CH15-OS8e.pdf

Thomas, E. M., Temko, A., Marnane, W. P., Boylan, G. B., & Lightbody, G. (2013). Discriminative and generative classification techniques applied to automated neonatal seizure detection. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2012.2237035

Truong, T.C; Diep, Q.B.; & Zelinka, I. (2020). *Artificial Intelligence in the Cyber Domain:* Offense and Defense. Symmetry 2020, 12, 410.

Umamaheswari, K., and Sujatha, S., (2017). *Impregnable Defence Architecture using Dynamic Correlation-based Graded Intrusion Detection System for Cloud, Defence Science* Journal, Vol. 67, No. 6, November 2017, pp. 645-653, DOI : 10.14429/dsj.67.11118.

Wilson, B. M. R., Khazaei, B., & Hirsch, L. (2015, November). *Enablers and barriers of cloud adoption among Small and Medium Enterprises in Tamil Nadu.* In: 2015 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) (pp. 140-145). IEEE.