Comparative Analysis of Knowledge Base System Development Approaches in the context of ethio telecom

BY: BIRUKTAIT FIKRE

ADVISER: BENEYAM BEREHANU (PHD)

A Thesis submitted to
School of Electrical and Computer Engineering
Addis Ababa Institute of Technology
Addis Ababa University

in partial fulfillment of the requirements for the degree of Master of Science in Telecommunication Engineering

Addis Ababa University
Addis Ababa, Ethiopia
January , 2022

Declaration

I, the undersigned, declare that the thesis comprises my own work in compliance with internationally accepted practices; I have fully acknowledged and referred all materials used in this thesis work.

Biruktait Fikre

Name

Signature

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

This is to certify that the thesis prepared by **Biruktait Fikre**, entitled *Comparative Analysis of Knowledge Base System Development Approaches in the context of ethio telecom* and submitted in partial fulfillment of the requirements for the degree of Master of Science Telecommunication Engineering complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Internal Examiner	______________	Signature	________	Date	_____
External Examiner	______________	Signature	________	Date	_____
Adviser	Beneyam Berehanu (PhD)	Signature	________	Date	_____

____

Dean, School of Electrical and Computer Engineering

ABSTRACT

For a successful business, telecom companies need to provide consistent, quality and timely maintenance and support to their customers. This is difficult to achieve without timely availability of domain experts that can efficiently be managed using the Knowledge Base System (KBS). Different approaches are applied to implement KBS and researchers study accuracy performance of the approaches. However, there are limited performance studies on the approaches in the context of telecom companies and none for ethio telecom. Most of the studies are also only focused to accuracy performance of the approaches and other relevant performance metrics that have an impact on the incident handling time of Information Technology (IT) experts are not addressed well.

This thesis work presents performance comparison for KBS development approaches in the context of operation support and maintenance process of ethio telecom. Rule-based, Case-based and Artificial Neural Network (ANN)-based KBS development approaches are considered and these approaches are modeled using data collected from ethio telecom technical support experts' knowledge and experience through interviews and data analysis. For the performance analysis, data preprocessing techniques including data reduction, data cleaning, data integration and data transformation are applied on the collected datasets. Execution time, accuracy, F-measure, precision, recall, error rate and mean square error metrics that have an impact on incident handling time are used for the performance evaluation. Python programming language is used to model, train, evaluate and compare the selected approaches.

Achieved performance results show that ANN-based KBS development approach has a better performance in terms of building and executing the model, taking a minimum time of 0.95 seconds. Furthermore, rule-based approach accomplishes the highest accuracy that is 98.6 %, with minimum error in identifying the class labels which are incident types. Based on the evaluation results on the execution time or response time, accuracy, and error measures which have an impact on the incident handling time, a hybrid KBS development approach from the ANN and rule-based KBS development approach can provide the best aggregate result.

KEYWORDS

KBS, Rule-based development approach, Case-based development approach, ANN-based development approach

ACKNOWLEDGMENTS

First of all I would like to thank the almighty God for giving me a good health, noble and supportive families, inner peace and strength to pass through all my challenges. Next to the almighty God, I would like to thank our lady of the saint virgin Mary for listening to my prayers and helping me in all things without leaving me since my childhood.

My special thanks goes to my adviser Beneyam Berehanu (PhD) for his time, advise and followup during this work. I would also like to thank my evaluaters Rosa tsegaye (PhD) and Ephrem Teshale(PhD) for their valuable feedback given on the progress session.

I would also like to thank my colleagues for their willingness to provide me the required data source for this work, to give me their time for interview and for their continuous support.

Finally yet importantly, I would like to give my special gratitude to my beloved families for their unconditional love, support and commitment. I want to say loudly that I'm lucky to have you, and you all are my strength.

To my newborn baby Mataria H/georgis, well came to my life. You can't imagine how God changed my life and blessed me by bringing you into my life. you have no idea how I feel about you, I can't express my happiness to have you too. I know I took your time to accomplish my thesis work, but since from now on I have a plentiful time to care you without any distractions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI	Artificial Intelligence
ANN	Artificial Neural Network
CBR	case based reasoning
CSD	customer service division
EOLC	Epistemic Ontology Language with Constraints
ES	Expert System
FAQ	Frequently asked question
GUI	Graphical User Interface
IT	Information Technology
IMARS	Incident Management and Reporting System
IS	Information Security
KA	Knowledge acquisition
KR	knowledge representation
KB	Knowledge Base
KBS	Knowledge Base System
KM	Knowledge Management
KMS	Knowledge Management System
LOOP	Logic and Object Oriented Programming
MSE	Mean Square Error
NN	Neural Network
RBR	rule based reasoning
TS	technical support
UI	User Interface

1 INTRODUCTION

The term knowledge is defined as a high-value form of information, which provides a competitive significance to a company and is mostly used for decision and action support [1, 2]. It is combined with company policy, strategy, rules, procedures, expert experience, interpretation, context, reflection, and skills. There are two basic types of knowledge. The first one is implicit or tacit knowledge, which resides in people as mental models, experience, and skills [1, 2]. The other one is explicit knowledge, which can be communicated and captured from formal models, rules, and procedures [1, 2].

From time to time knowledge is gradually taken as an organization's valuable resource and becoming a key to organizational success [3],[4]. In order to have a successive journey and drive as competitive, companies are required to manage their organization's knowledge [3, 4].

Knowledge Management (KM) is a collection of formal processes that direct on how a piece of knowledge is captured, created, distilled, organized, stored, disseminated, utilized, and retained throughout the organization [2]-[4]. Managing both, implicit and explicit knowledge of a company provide enormous benefits [1]-[4]. Such as

- Filling knowledge gaps among the same section employees by creating expert knowledge sharing means
- Accelerate self-support for service or product users
- Increase customer satisfaction by eliminating time wasted to communicate or ask domain experts and by minimizing request or incident handling time
- Accomplish learning without diverting experienced employees and without spending time and money on training

- Helps to accomplish employees tasks by providing updated information, by enabling to perform the right procedure and model, by enabling to follow the right rules and regulations
- Offer knowledge for innovation

The KM process by itself is not technical. IT emerges a computer system with an intelligent technical means to manage an organization's knowledge; known as a Knowledge Management System (KMS), KBS, or Expert System (ES) [1]-[5], [6].

KBS and ES used interchangeably, to express Artificial Intelligence (AI) based system designed to solve a problem using expert's decision-making procedure and ability by simulating knowledge, methods, and experience of experts [1]-[7].

The value of a KMS or KBS is recognized by different organizations and applied to problems in so many fields. From researches, fields like engineering, electronics, law, agriculture, medicine, manufacturing, metrology, geology, physics, forestry, finance, IT, and telecommunication are mostly mentioned fields, which implement KMS [3] -[7].

In the telecommunication field, there are a number of problems that need to implement KBS as a solution. Within telecom companies' environment, there exist numerous network equipment, applications, systems, software, and hardware used for providing telecom service and products. For a successful business telecom companies need to sustain a good quality of service by providing consistent and timely maintenance and support for their customers. This is succeeded by hiring and training IT experts. But to maintain telecom services, to handle incidents that may happen on the service and equipment, and to provide timely support for customer requests, IT experts need to be timely available. If IT experts can't be available due to different reasons, telecom companies encounter difficulties to provide a sustainable quality of service. However, this can efficiently be managed using KBS by collecting, storing, sharing, and managing IT Expert's knowledge and experience on solving problems and decision-making ability.

Researchers propose different approaches like rule-based, semantic networks or ontology-based, case-based, ANN Based, frame-based, object-oriented based and scripts based approaches to implement KBS or ES [7]- [8]. Each approach has its own advantages, disadvantages and performance. Finding a suitable approach, to build and implement the best KBS for a specific company, is not an easy task and needs deep investigation.

1.1 BACKGROUND OF ETHIO TELECOM

Ethio telecom is a fully government-owned sole provider of telecom service company in Ethiopia since 1894. The company tries to increase the number of its services and subscribers in each fiscal year from its establishment. Ethio telecom, as a telecom service provider, implements the latest technologies to provide modern service for its enterprise and individual customers. The company also continues to build and expand its network infrastructure, implement applications and systems that can enable the company to offer recent services and technologies in a good quality of services. Because of different causes the employed network infrastructure, applications, systems, software, hardware, and equipment may face different problems, which are called incidents in the company trend, and affect the functionality of the company. This will affect the company employees from achieving their dedicated duty on the company's mission and on serving customers. It also affects customers from getting subscribed service with good quality.

In order to provide technical support and maintenance by handling such kinds of service and operation interruption, ethio telecom organizes the operation and maintenance department in its structure. Those departments are incorporated within each division to support the company employees as internal customers and service users as external customers. Some of the operation and maintenance departments are specifically in charge of supporting the company divisions and departments to have the required system and applications privilege, to resolve their technical difficulties, and to maintain their resources and applications based on their scope. Those divisions and departments, that need support from operation and maintenance, serve and support external customers. Though supporting these divisions and departments is directly related with supporting external customers, hence problems that happened internally affect external customers.

Ethio telecom hired IT technicians who can satisfy the operation and maintenance department requirements. Employees assigned to operation and maintenance departments get induction and on-the-job training on the way of handling expected customer requests and maintaining resources, services, and applications. The company also provides additional training and workshops to expert its IT technicians. Employees working in operation and maintenance departments by themselves try to acquire and update their knowledge and ability of operation support and maintenance through education and experience with personal effort and using the company's training opportunity. They continuously keep going to improve themselves in incident handling, maintenance, and technical support and bring themselves up to specialist and expert levels. Hence IT experts use their knowledge and experience to solve the problems. Handling incidents and requests require employees to have a better and fast way of responding. This makes the time required to handle the problem depend on the expert's ability.

Those technical support specialists and experts may assigned to another duty or transfer to other positions or resign from the company or may take a longer leave. Such types of cases forced the support departments to handle customer problems with the remaining staff, which may not have enough knowledge on incident handling, or with newly hired ones. This makes the company to face a challenge in providing consistent operation support by handling the same incidents and customer requests with a lower capability as compared to its previous capability. To sustain the previous way of incident handling and to obtain specialist and expert trends of handling, support staff will try to contact experts through different means of communication. Specialists and experts may not be available at the required time or explain the solution explicitly; which may take extra time for understanding. This will result in handling the same problem with much time, which causes an increment in a customer complaint, decrease customer satisfaction, and avert the support and the company from meeting its target achievement. As a solution, from the previous and current trend, the support department facilitates additional training. This will cause the company to spend extra expenditure and needs additional time to prepare, schedule, and perform the training. These all the mentioned solutions are not time and cost-effective.

In order to solve this problem, there must be a means of managing and sharing domain experts, which are specialists and experts, knowledge and experience on incident handling among the operation support employees. For this purpose, implementing KBS is an efficient way. To do so, there are several KBS development approaches and techniques, that help to manage knowledge and facilitate expert knowledge sharing, proposed by researchers. The proposed KBS development approaches include rule-based, semantic networks or ontology-based, case-based, ANN Based, frame-based, object-oriented based and scripts-based approaches [7]-[8]; and have different capabilities and performance.

Studying the performance of the available KBS development approaches is required to identify the best approach for implementing a better KBS which is suitable for ethio telecom.

1.2 STATEMENT OF THE PROBLEM

Telecom companies, including Ethiotelecom, must provide consistent support and maintenance for their customers to achieve a successful business. This in turn, required having domain experts that can have enough knowledge and experience in incident handling and avail in the required time. Domain experts may not be timely available due to different reasons; remaining or newly hired staff may cannot able to handle incidents as domain experts do. Under this circumstance, providing consistent and quality operation maintenance and support cannot be achieved. Such kinds of problems result in increasing customer complaints, decreasing customer satisfaction, and prevent the company from meeting its targeted goal and missions. To overcome this challenge solutions like trying to contact domain experts through different communication means and facilitating additional training for the remaining and new staff are not time and cost-effective. This can be resolved efficiently by using KBS to manage and share domain experts knowledge and experience on incident handling among the operation support employees.

There are different approaches that can be applied to implement KBS. Those approaches have different capabilities and performance. Studying the performance of the available KBS development approaches is needed to select the appropriate one. Yet there are limited KBS development approaches performance studies in the context of telecom companies and none for ethio telecom. Most of the studies are also focused only on the accuracy performance of the approach and other relevant performance metrics that have an impact on incident handling are not addressed well.

There is a need for a study that can address such research gaps and difficulties in identifying a better KBS development approach which is suitable for ethio telecom to solve the mentioned problem. This research focuses on proposing a better KBS development approach suitable for ethio telecom by studying and comparing KBS development approach performance that has an impact on incident handling other than accuracy.

1.3 OBJECTIVE

1.3.1 *General Objective*

The main objective of this study is to compare knowledge base system development approaches and propose a better approach suitable to manage experts' knowledge in the ethio telecom operation and maintenance process.

1.3.2 *Specific Objectives*

The specific objectives of this study are:

- To identify and understand different KBS development approaches
- To build and train selected approaches used for KBS development
- To test and analyses the performance of selected KBS development approaches
- To compare selected approaches used for KBS development

- To propose the better approaches suitable to develop KBS for ethio telecom operation and maintenance process

1.4 METHODOLOGY

In order to achieve the general and specific objectives of this study, the following methodologies and techniques are applied.

1.4.1 *Literature Review*

Relevant literature and articles including conferences, journals, thesis, and books related to the Knowledge Base system are reviewed and grasp a better understanding of:

- Knowledge Base system
- Approaches, techniques used to model or develop a KBS and select the appropriate approach
- Tools used to implement KBS and select the appropriate tool to implement the selected knowledge base system model
- Data required to model and implement KBS

1.4.2 *Data Collection*

Both primary and secondary sources of data were used to acquire the required knowledge for this study. Knowledge is mainly gathered from Ethio-Telecom's technical support (TS) section. TS structurally found in CSD operation support department and responsible to support customer service employees to accomplish their duty and to support external customers.

As a primary data source, domain knowledge is collected through interviews with domain experts that spent in the technical support section of ethio telecom to gather the cause and possible solutions for selected incident types.

Although many sampling techniques are available to conduct research, the purposive sampling technique is applied for this study to select domain expert [9]. Purposive sampling is defined as " judgmental sampling that makes the conscious selection of certain subject or element to include in the study " .

The basic criteria considered to select domain experts are their previous and or current position, educational background, and year of experience. Based on those criteria five (5) domain experts are selected from the technical support section of ethio telecom using purposive sampling.

Unstructured interview, which is sometimes referred to as discovery or informal interviews, the technique also applied to acquire the required knowledge [9, 10]. In this interview technique, the interviewer inquires open-ended questions related to the research topic instead of using a formal set of questions [10]. This nature of the interview technique enables the interviewer to modify or change the order of the question based on the response or experience of the interview candidate [10]. It also makes the interview candidate to be flexible in sharing experiences like having a natural informal conversation [10]. Therefore, using the unstructured interview technique knowledge is acquired from domain experts by focusing on the expert's experiences and documents, if any, used to handle the selected incident types and customer requests.

As a secondary data source, 6 months recorded incident log and report is extracted and document analysis is applied: by referring technical support written documents, manuals, and incident log and report as a baseline, additional relevant literature also reviewed.

1.4.3 *Data Preprocessing*

The collected expert knowledge is preprocessed by reducing, cleaning, transforming, and integrating the dataset, to have a clean, complete, and suitable dataset for the selected knowledge base system development approach models and tools.

1.4.4 *Building Knowledge Base System Model*

Building a well-formed knowledge base system is required to select an appropriate technique to structure the knowledge.

From the literature review three mostly applicable knowledge base system development approaches are selected to be modeled, evaluated and compared based on their performance. Those are rule-based, case-based, and ANN-based development approaches.

The selected three KBS development approaches models are built, as presented in chapter four, according to the required knowledge representation.

1.4.5 *Implementation Tool*

There are numerous programming languages established to implement Knowledge base systems. Researchers usually used programming languages like PROLOG, CLIPS, LISP, Logic and Object Oriented Programming (LOOP), OPS5, Epistemic Ontology Language with Constraints (EOLC), python, mathlab , ARCHLOG , ADA, JAVA and C++ [11], [12], [13], [14].

Among those programming languages python is selected to be used as a tool for building the selected KBS development approaches models. It is selected because of its portability, applicability, increased productivity, easy and fast debugging capability, robustness, full development environment, and sustainable performance [11] - [14].

Python is one of an open source software designed based on object oriented programming concept. It is more applicable to implement knowledge base systems or expert systems; hence, it is a widespread programming language for data science [14],[15], [16].

1.4.6 *Evaluation and Result Discussion*

To evaluate the performance of the modeled knowledge base development approaches evaluation metrics such as execution time, accuracy, F-measure, preci-

sion, recall, error rates and mean square error are selected.
The performance of the KBS model is evaluated and compared based on selected evaluation metrics with the involvement of domain experts. Based on the evaluation and comparison result, the KBS development approach with better performance is proposed for ethio telecom.

1.5 SCOPE AND LIMITATIONS OF THE STUDY

This study aims on comparing the performance of rule-based, case-based, and ANN based knowledge base system development approaches.

Although there are different KBS development approaches this research limited to study rule-based, case-based, and ANN approaches. In addition, though there exist a number of incident types, the study focuses on the top ten incident types.

1.6 CONTRIBUTIONS OF THE RESEARCH

Ethio telecom is beneficiary from this study because it identifies which knowledge base development approach is appropriate for the company to build the best knowledge base system. In addition, if the proposed solution is implemented it enhance the operation and maintenance support capability of telecom companies, like ethio telecom. By enabling companies

- To have a fast and accurate KBS with minor errors in identifying the occurred incident types and recommend expert solutions.
- To overcome challenges that occurred when domain experts were unable to exist at the required time.
- To provide consistent and smooth operation support and maintenance with available IT technicians by adopting expert solutions.
- Technicians, who have not enough experience in the domain area, to solve incidents easily and accurately in a minimum time using experts' knowledge.

- To minimize investment run to train new technicians by enabling experts' knowledge sharing.
- To increase customer satisfaction, minimize customer complaints, and increase company revenue by minimizing service interruption time.

From the literature review, it is clearly shown that there is no specific work done on comparing the three knowledge-based development approaches: rule-based, case-based, and ANN-based approaches using additional evaluation metrics other than accuracy. In addition, there are limited researches done related to KBS development approaches applicability for telecom sectors. So, the result of this study can be used as an input for the research areas community.

1.7 LITERATURE REVIEW

Many researchers conducted to provide approaches, which may have different reasoning techniques and help to build a knowledge base system or expert system. Among such researches, some related work is reviewed and presented below.

L. Atymtayeva, K. Kozhakhmet, and G. Bortsova [6] developed an approach used to build a Knowledge Base (KB) for Information Security (IS) expert system based on ISO 2700k IS standard. The constructed ontology KB consists of assets (objects that need to be protected), sources (standards), threats that could influence assets, vulnerabilities that can exploit those threats, steps (guideline or controls that might mitigate those vulnerabilities), and relationships among them from IS expert's knowledge. The developed approach recommends appropriate measures that can be taken to avoid a risk that might happen in the organizations and improve information system security level by analyzing the threats or vulnerability types with IS standards. The developed model serves as IS audit process, which avoids previous security audit process expense in time, cost, and human resource involvement. The result enables IS management to discover and decide on the right IS policy by reducing the cost spent for IS audit.

Latha B. Kaimal, Abhir Raj Metkar, and Rakesh G [17] present an enhanced algorithm suitable to the real-time expert system used for alarm's root cause analysis within a power plant. The researchers compared the available association-based machine learning algorithms and select an efficient algorithm, which uses real data exist in a power plant domain and generates all significant rules. The authors develop the expert system using programming languages Java, Flex, and MySql Database. From available association-based learning algorithms, such as Apriori, predictive Apriori, FP growth, and Tertius, the authors found that the Tertius algorithm is appropriate for this application. The result shows that the learning engine that uses an association rule-based algorithm enhances the real-time expert system by supporting the establishment of new rules.

Sourav Mandal, Sumanta Chatterjee, and Biswarup Neogi [18] introduced an intelligent expert system which helps to diagnose and troubleshoot computer hardware faults based on expert system and AI. The designed system receives computer system defects, diagnoses the defect by applying formal fuzzy-logic rule-based reasoning from its intelligent KB, and recommends a probable solution to resolve the problem. The authors present the extension of their previous work, implemented with Turbo Prolog programming language and database management system, by using Object Oriented technology and web-based programming approach in addition to exploring the previous expert system. Even if the system is not tested, the authors put their expectation of the system's ability in helping to repair computer system faults and save time and cost by suggesting appropriate time-consuming maintenance tasks.

Mazlina Md Mustaffa, A.S.Shibghatullah, A.S.H.Basari, and B.Hussin [19] presented an intelligent expert system that helps for diagnosing computer hardware failures and making a decision on a solution. The system records the computer hardware problem, symptoms, and domain knowledge expert solutions; then it diagnoses the computer hardware failure using rule-based technique and presents possible causes with appropriate solutions. The authors use a formal software development process to develop the proposed system. Following the process they mentioned that the expert system was evaluated by human experts but they did not present the result, they only put their general expectation. In general, the proposed system expect to assist technicians in making a quick decision in diagnosing computer hardware failure and providing optimum solutions. It also expects to solve the problems in diagnosing computer hardware failure, their causes, and solutions. Such as the trial and time were taken to actually detect and identify the cause of computer hardware failure.

Yingying Wang, Qiuju Li, Chang, Hongwei Chen, and Guohua Zang [20] introduce fault diagnosis expert system by using and combining advantages of Neural Network (NN) and fault tree technologies. Researchers use control box faults taken from the digital control system to simulate, train and test the designed expert system model. The result shows that the proposed model is more accurate and fast in fault diagnosis by enabling the system to locate the source of failure.

Finally, the reviewed related work presents an expert system suitable for the selected domain. They used different approaches to acquire, represent, model, and implement KBS. They also use different mechanisms like an expert test to evaluate the accuracy of their system. Even if there are additional metrics that have an impact on the incident handling time of IT technicians, none of them compares their proposed approach with another similar purpose KBS development approach to evaluate its performance other than its accuracy. In addition to this, the performance of such development approaches is not studied in the context of telecom companies, especially ethio telecom.

1.8 THESIS ORGANIZATION

This study is organized into six chapters. Chapter 1 is the introduction part which contains the background of the study, statement of the problem, objectives, methodology, scope and limitation of the study, the contribution of the study, and related literature reviews. The second chapter Chapter 2 provide detailed information about the knowledge base system overview including its background and architecture with its component description. The overall discussion of the KBS development approaches ,which are selected for this study cover under Chapter 3. Tasks performed under the process of modeling the selected knowledge base system development approaches stated in Chapter 4. Chapter 5 present the result found in the performance evaluation of the development approaches. Chapter 6 discuss the conclusion of this study based on the obtained experiment result and recommendations for future work.

2

KNOWLEDGE BASED SYSTEM

2.1 INTRODUCTION

The knowledge-based system also known as an expert system, is an artificial intelligence-based system designed to solve a problem using an expert's decision-making procedure and ability by simulating knowledge, methods, and experience of experts [1] - [6].

An artificial intelligence by itself is devised in 1956 and denote the latest science that makes computer programs intelligent to do things by adopting the perceiving, understanding, manipulation, predicting ability of the human being [5, 21, 22]. This makes AI entities to handle many problems in the same way as human beings handle it [5, 21, 22]. Starting from its emerging time, which is 1956, AI became a scientific basis for different technology fields including KBS.

Knowledge-based system uses human expertise domain knowledge, stored in an intelligent computer program, called knowledge base, then drive expert knowledge from it when users' needs access to the knowledge. It can consult by giving a suggestion or advice like human experts on how anyone can solve a problem. KBS enables to save money, time, and effort by making several experts' knowledge to be available in anywhere and anytime without the need of the experts' physical existence. It also smoothens the consistency of work by providing experts' knowledge and making anyone to function in a productive manner at a higher level as an expert.

KBS provide numerous benefits by overcoming domain expert's limitation such as [4, 23]:

- There may be scarcity in domain experts.

- Experts' may take a longer leave, assign to another duty, resign or die
- Experts' may forget details in the problem-solving process or procedure
- Experts' may get physically or mentally tired from workload
- Experts' may not be consistent in their day-to-day decisions
- Experts' are limited in place and time
- Some experts' may not need to share their experiences, so they may lie about or hide their experience

2.2 ARCHITECTURE OF KNOWLEDGE-BASED SYSTEM

KBS as a system consists of an interconnected system component with their own roles working together to provide the required functionality. As shown in the architecture of KBS presented in fig 2.1, the essential components of the KBS are knowledge acquisition module, knowledge representation module, KB, inference engine, and User Interface (UI) [7], [8]. Each of these components with their roles is briefly explained in the following sections.

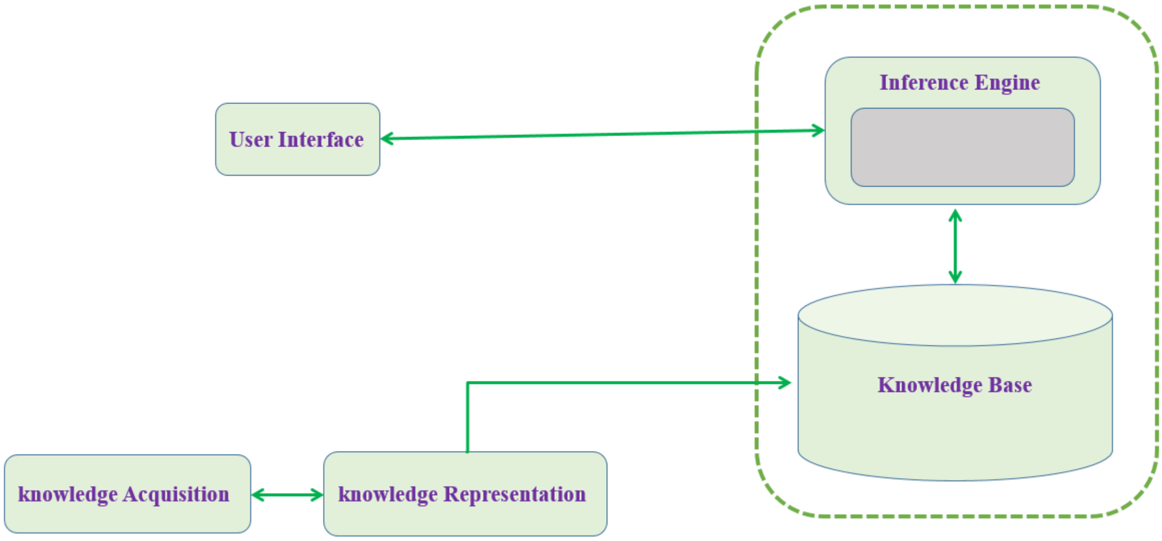


Figure 2.1: Architecture of a Knowledge-Based System [8]

2.2.1 *Knowledge Acquisition*

The first step in building KBS is Knowledge acquisition (KA). KA is the process of extracting relevant knowledge from specified data sources such as domain experts, documents like books, manuals, and journals [7] – [8]. KA extract knowledge to accumulate and transfer experts' knowledge and problem-solving skill into computer programs and system users. Potential data sources of knowledge include domain experts', domain-specific literature and researches, textbooks, databases, and the like. Techniques such as interviews, data analysis, questioners, and observations can be used to acquire knowledge from potential data sources. KA includes eliciting and structuring of knowledge obtained from different data sources [7] – [8].

The KA process determines the performance, usefulness, and trustworthiness of KBS. So, more attention must be given to the effectiveness of the techniques used to acquire reliable, accurate, and valid knowledge [7]- [8].

In order to build the KBS efficiently, knowledge engineer must carry out the following steps sequentially during KA process [8].

- Elicit the data obtained from domain experts and document analysis
- Interpret the acquired knowledge and conclude domain expert knowledge and reasoning process
- Construct the model by structuring the expert's knowledge
- Repeat step I-IV until the KBS became fully functional

2.2.2 *Knowledge Representation*

After extracting the required domain knowledge, knowledge must be encoded in a systematic means [7]- [8]. This implies a knowledge representation (KR). KR includes preparing a knowledge map and the documenting, structuring, encoding of acquired knowledge. KR is the process of transforming and encoding the acquired knowledge into a systematic means or computer understandable form to make the knowledge suitable for the selected KBS development approaches [7] –

[8]. Through the process of knowledge representation knowledge engineer makes the knowledge to be suitable for the designed knowledge base. KR differ based on the selected KBS development approaches. Different KBS development approaches have different ways of knowledge representation technique.

2.2.3 *Knowledge Base*

The body of KBS mainly consists of the inference engine and KB, which exist as separate modules and work together closely [7]- [8]. They provide the main function of KBS.

The KB is a repository that stores all relevant knowledge acquired from domain experts through knowledge acquisition process [7]- [8]. The acquired knowledge, which is structured and encoded in the knowledge representation process, is stored in the knowledge base. The structure of the knowledge base differs based on the selected KBS development approaches. Different KBS development approaches have different ways of structuring the knowledge in the knowledge base.

2.2.4 *Inference Engine*

The inference engine is an instruction on how the knowledge stored in knowledge-based is used for problem-solving [7]- [8]. It performs a search and deduces facts or draws conclusions from expert knowledge stored in the knowledge base. Inference engines accept end-user questions in a natural language form, formulate them to be understandable for a computer program, find the solution for the end-user questions and present the answer in the form of natural language. Inference engine uses reasoning mechanisms to manipulate, control, uses, and interpret expert knowledge and provides expert recommendations or advice [7]- [8].

2.2.5 *User Interface*

Once the construction of the KBS is completed, it requires one critical module called the user interface to interact with system users simply by hiding KBS's complex internal structure [7]- [8]. The quality of UI may determine the acceptance of KBS. In order to get user acceptance, the KBS UI must be user-friendly. KBS uses a Graphical User Interface (GUI) or natural language base question and answer, menu-driven user styles to implement the UI [7]- [8]. However, currently, most KBS use GUI. The user interface is where the KBS accepts end-user questions as an input and display experts' solution for the end-user as an output.

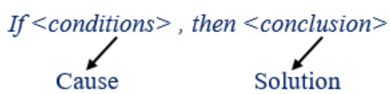
3

KNOWLEDGE BASED SYSTEM DEVELOPMENT APPROACHES

Knowledge base system development approach, also known as knowledge base reasoning technique, is used to implement KBS or ES [7]- [8]. Knowledge base system development approaches differ on the way they are used to represent knowledge, on their knowledge base structure design, and their inference engine reasoning. The well-known Knowledge base system development approaches are rule-based, semantic networks or ontology-based, case-based, ANN Based, frame-based, object-oriented based and scripts based approaches [7]- [8]. For the purpose of this study, from the available reasoning mechanisms, the most commonly used three reasoning techniques; rule-based, case-based, and ANN-based techniques are selected and discussed as follows.

3.1 RULE BASED DEVELOPMENT APPROACHES

Rule-based knowledge base system development approach, which is also known as rule based reasoning (RBR) technique is one of the most popular reasoning techniques that represents a piece of knowledge in the form of rules facts [24]. The rule is a conditional statement that derives a conclusion based on the fulfillment of a condition. The basic form of rule-based reasoning is the "IF-THEN " approach.



From the "IF-THEN" statement, the "IF" side expresses the premises or condition that needs to be true in order to execute the rule, and the "THEN" side of the statement states action needs to be taken or the conclusion needs to derive.

In the rule-based KBS development approach two inference methods are available. They are known as forward-chaining (Data-Driven) and Backward chaining (Goal Driven) [25]. Both inference methods follow the same procedure and have the same function, they only differ in their inference technique and in the structure of the KB.

Forward chaining (Data-Driven)
In the forward chaining inference method, the inference engine receives a problem as a condition and derives the conclusion that satisfies the condition. First, the system will crosscheck all the rules to find the one whose condition satisfies. So, with this scenario, a rule is executed at least once whether a fact that satisfied it is available in the database or not.

Backward chaining (Goal Driven)
This inference method only focuses on the rules that are applicable to a particular goal or conclusion. Unlike to forward-chaining, backward chaining does not check rules that are not related to its conclusion or goal. Backward chaining receives a problem as a conclusion and tries to find premises or conditions that will cause the conclusion. In another expression, the system accepts a goal and then tries to find the fact that can match with the goal.

3.2 CASE BASED DEVELOPMENT APPROACHES

Case-based knowledge base system development approach, also known as case based reasoning (CBR) technique, is an automated reasoning and decision process whereby it solves new problems through experts' experience accumulated in solving previous ones [26, 27]. CBR is built by storing the condition of the problem and the action taken at that time then the KBS focused on recognizing the similarity of new problems to existing previous ones. It is used for diagnosis, prediction, classification, or recommendation of a solution to the new case by remembering previous similar case[26, 27].

CBR represent a piece of knowledge in the form of cases. Cases, in CBR, can be taken as a previously experienced situation collected and stored over time. CBR focuses on managing cases to solve a new case by finding a previously recorded situation that is similar with the new one. CBR structured as a four-step process named as four R's which are: Retrieval, Reuse, Revision, and Retention. The below figure represents four R's as a case-based approach cycle. Each process is also stated in detail in the following section.

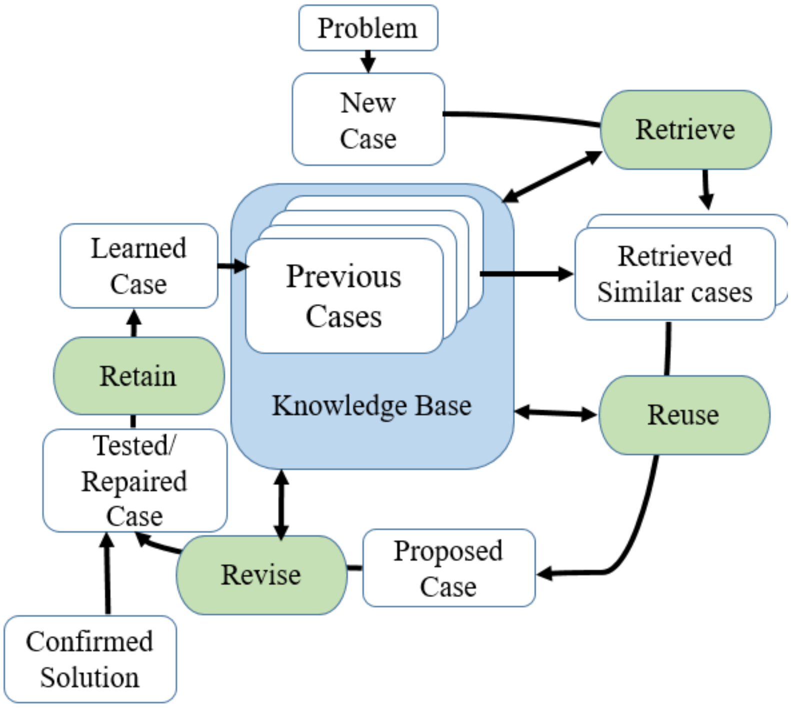


Figure 3.1: Case based cycle [26]

Retrieval is the process of finding a case similar to the current situation or state.

: It relies on ability to assess any two states based on their applicability.

: Adjustable weight vector to control direction of the hyperplane

: The retrial process uses a similarity measure to measure how much the previous cases are similar to the new case.

: **Similarity Measure :** also known as case matching

: match new case with the previous case from the case based to find the solution.

: Select cases that have the same or near the same solution and can be adopted easily to the current problem.

: uses attribute weight, which is assigned based on their importance, to calculate degree of similarity.

: There are two types of similarities : -

Local and Global similarity

: Local similarity used to compute the similarity between new case attribute and previous case attributes value at feature level.

: Attributes value can be discrete and numeric or continuous.

: Global similarity is a weight sum of local similarity computed at case or object level

Reuse : is a process when the system retrieve a case and propose it as a valid action to apply for the current state.

Revision : is a process to evaluate through a serious of metrics or simulation on how well the proposed solution will perform and whether it is practical to apply it for the new case.

Retention : is a process of storing the result of the new or previous case experience in memory to be used for future.

In order to build the KBS using the case-based development approach, knowledge must be acquired from previous situations and experts' solving experience first, then knowledge must be represented in the form of cases. The knowledge base and the inference engine should be designed based on the case-based cycle process. The user interface should also align with and support the case-based process.

3.3 ANN BASED DEVELOPMENT APPROACH

In the ANN based development approach, also known as ANN based reasoning, knowledge is represented as a structure of human brain systems [28, 29]. It is designed by making artificial neurons to be an approximately simulated model of human brain neurons. ANN has a large connected network, formed by neurons which takes an input and computes the desired output by processing the input values through multiple layers. ANN store the knowledge in the form of weight in which the value is decided based on the importance that particular input has to calculate the desired output. Some computation is taken place between input and output layers, in the hidden layers [28–30].

Layers are used to hold neurons and pass them to the subsequent layers. In the ANN based development approach neuron is an input point. Each neuron is a mathematical operation that takes its input to multiply it by the weight associated with it and passes the sum through the activation function to the other neurons.

As can be seen in fig 3.2 below ANN generally has three layers known as input layer, hidden layer, and output layer.

Input Layer:	accepts all the inputs provided by the user
Hidden Layers:	are set of layers that can be exist
	between input and output layers.
	: perform some computations which results in the output.
	: results are gained from this layer.
Output Layer:	is the layer where the output is delivered.

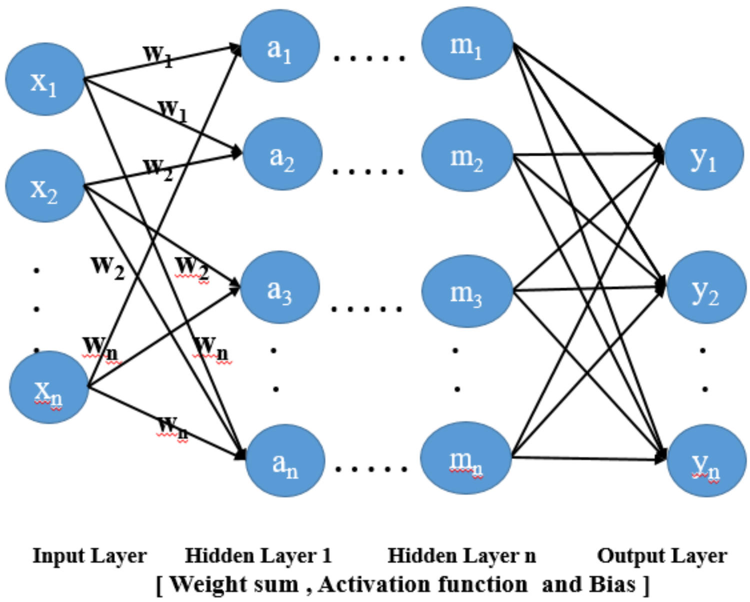


Figure 3.2: Artificial Neural Network (ANN) model [30]

In each layer, there are a number of neurons that take the input and perform a mathematical operation by multiplying the input with weights associated with it and passing the sum through the activation function to the other neurons. The activation function computes the output from the weight of some of the inputs. It also normalizes the computed input to produce an output. Hence ANN deals with numeric data knowledge, knowledge needs to be converted from text to binary.

There are a number of ANN types. The most applicable ANN types are feed-forward neural network, convolutional neural network, and recurrent neural network.

Feedforward Neural Network (Artificial Neuron)

In this ANN type, the input in the NN travels in one direction. It only used front propagation wave and backpropagation is not allowed in this ANN type. Feed-forward NN is the simplest form of ANN, which makes it easier to be build and maintained. The input data is entered through the input layer nodes, the weighted sum of the nodes is calculated using classifying activation function and the output is exited on the output layer nodes. Such type of ANN may or may not have a hidden layer [31] - [34]. Feedforward neural network applications are found in speech recognition, computer vision, face recognition, and simple classification [31] - [34].

Convolutional Neural Network

A convolutional neural network encloses a three-dimensional array to arrange neurons, instead of having a two-dimensional array as standard. The first layer of a convolutional neural network is known as the convolutional layer. In the convolutional layer, neurons can only process from the small part of the input features by filtering them in batch-wise and computing this operation many times until the full visual image or the input is processed. The convolutional neural network is similar to a feedforward network, yet its neurons have learnable weight. Convolutional Neural Network applications are found in speech recognition, computer vision, video recognition, machine translation, semantic parsing, signal processing, image classification, paraphrase detection, and image processing[31] - [34].

Recurrent Neural Network

In a Recurrent Neural Network, the output of a layer is saved in its neurons and feedback to the input in order to predict the outcome of a layer. The recurrent computational process is accomplished by making neurons to retain and remember information from previous steps. In the forward propagation stage of the recurrent neural network, input information is executed using a recurrent computational process. In the backward propagation, a layer outcome is predicted using previously stored information in neurons. If the prediction is wrong, error correction or learning rate means is used to make small changes that make recurrent neural networks to perform the right prediction. Recurrent neural networks applications are found in the text to speech processing, text processing, image tagger, translation, and semantic analysis[31] - [34].

4

EXPERIMENTAL ANALYSIS

The overall and detailed experimental process performed to conduct this thesis work is specified in this chapter. The overall experimental process of this research work is shown in fig 4.1 below. As illustrated in figure 4.1 the experimental process is composed of four components such as data collection, data preprocessing, modeling or building knowledge base development approaches, and finally model evaluation and comparison. The detailed activities carried out under each step of the experimental process are described in the next section of this chapter.

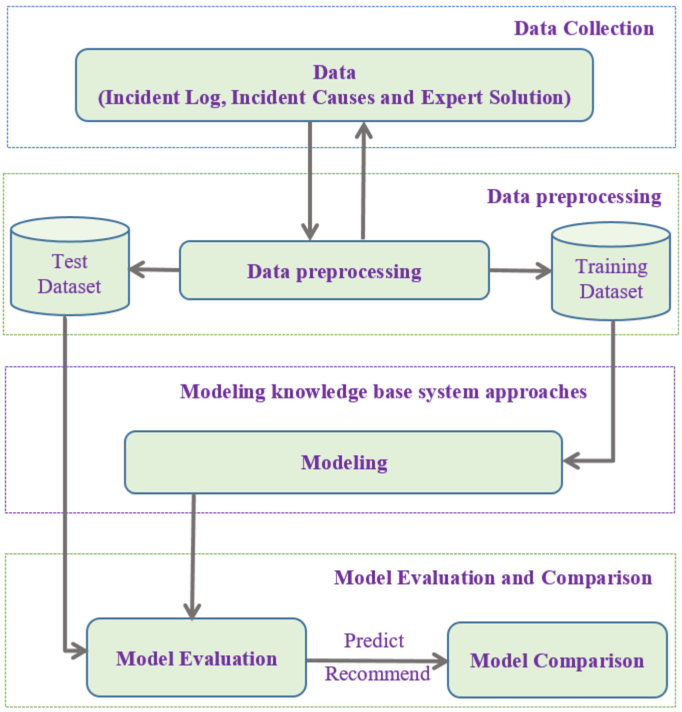


Figure 4.1: Overall Experimental Process [35]

4.1 DATA COLLECTION

For this study, both primary and secondary sources of data are used to collect the required data and to acquire the required knowledge from it. Interview with domain experts is used as a primary data source whereas as a secondary data source domain knowledge is collected using document analysis: by referring technical support written documents, manuals, incident logs, and incident reports.

Any problem that might happen in the network infrastructure, applications, systems, software, hardware, resource, and equipment is known as incidents in the ethio telecom trend. The main data used for this study is a collection of incidents that happened in the customer service division (CSD), which is stored in a system internally developed by Ethio-Telecom's technical support team and known as Incident Management and Reporting System (IMARS).

Each and every incident occurred in the CSD logged by the TS team in the IMARS. The team TS is the one who is responsible for providing operation and maintenance support for CSD employees to accomplish their duty and for supporting external customers.

For this work, the 6-month incident log extended from 01-09-2020 up to 28-02-2021 that contains 4531 incidents extracted from IMARS. The collected data consists of around 19 attributes. The attributes of the incident log with their description are illustrated in table 4.1.

No	Attribute	Description
1	Site	A Place where the room is located
2	Section	The name of the section
3	Room	Specific room where the incident happened
4	Owner	Specific name of the incident handler
5	Incident Type	General Category of the incident happened
6	Incident Description	The specific type of incident
7	Handler	Handler of the incident position
8	SD Number	Number of the incident if the incident is escalated to other section
9	Status	Status of the incident whether it is closed or opened
10	Raised Time	The time where the incident is happened
11	Reported Time	The time where the incident is reported to the handler
12	End Time	The time where the incident is handled
13	Frequency	Redundancy of the incident occurrence
14	Duration	For how long the incident is stayed
15	Remark	Description if any
16	Coach Group	Specific team if the incident is happened in Call center
17	Impacted User	The number of affected users by the incident
18	Created Date	The date reported to other section handler if the incident is escalated
19	Modified Date	Date where the status of the incident is updated

Table 4.1: Raw Data from Incident log

Not all the above incidents, incident types, and attributes are necessary for this work. So attributes, features, and the top 10 incident types are selected in the data preprocessing stage.

However, the incident log doesn't contain the causes or indicators of the incident, which are used to identify the incident types with their possible expert solution and mainly required for this study. Due to this reason, the remaining required data is needed to be collected through interviews with domain experts. In order to simplify the work, the interview is done after data preprocessing with selected top 10 incident types. The interview is held with domain experts, selected current and former TS employees, to identify the causes or indicators of the selected top 10 incident types and the expert's solutions for them.

Five (5) Domain experts were selected using purposive sampling techniques, although many sampling techniques are available to conduct research [9]. The basic criteria considered to select domain experts are their previous and or current position within the TS section, educational background, and year of experience in the TS section.

To perceive the required knowledge unstructured interview technique is applied. The interview with domain experts focuses on their expert's experiences and documents if any, used to identify each incident type and techniques used to handle the selected incidents. The interview questions cover on identifying the main probable causes that indicate the specific incident type and on how the domain experts handle the selected incident or their recommended solution for a specific incident. During the interview session, the collected data from each domain expert was recorded on an excel sheet. The selected and participated domain experts' profiles are presented in Table 4.2 below.

No	Current Position	Previous Position	Year of experience	Level of Education
1	IT Application Support Supervisor	IT Application Support Specialist	8	MSC.
2	IT Application Support Supervisor	IT Application Support Specialist	8	MSC.
3	IT Application Support Specialist	IT Application Admin	8	MSC.
4	IT Application Support Specialist	IT Application Admin	6	BSC.
5	IT Application Support Specialist	IT Application Admin	5	BSC.

Table 4.2: Profile of domain experts

In addition to the interview, data analysis on documents such as Frequently asked question (FAQ) prepared in TS and training documents guided to maintain implemented systems are done to assess causes or indicators used to identify the selected top ten incident types and their recommended expert solution.

From the result of unstructured interview and data analysis, 30 causes or indicators that are used to identify the selected top ten incident types are collected and listed in the table 4.3 below:

No	Incident cause (indicator)	No	Incident cause (Indicator)
1	Static IP assignment Problem	16	Corrupted MS Office application
2	DHCP Problem	17	OS Problem
3	Shortage of Static IP	18	Mcafee antivirus issue
4	Compatibility issues on browser	19	System stack
5	Network/ Intranet problem	20	IPCC system become busy
6	Mail size exceed	21	Multiple browser open at time
7	User account Problem	22	IE crushed or stack
8	Outlook set to offline	23	Outlook problem
9	IPCC CRM Integration failure	24	Microsoft Office License
10	Network interruption between TC and server side	25	PC performance problem
11	Cookies and temp file	26	Outlook does not closed properly
12	PC miss policy update	27	Attachment file size is beyond the limit
13	Open eye does not get a license	28	File corrupted or attacked by virus
14	MS Office activation problem	29	Application Software Corrupted
15	Un licensed MS Office	30	Application Software is not compatible for the file

Table 4.3: Incident causes (indicators) from Interview

4.2 DATA PREPROCESSING AND ATTRIBUTE SELECTION

Raw data may be inconsistent, incomplete and noisy [35],[36]. So preprocessing the data is essential. Data preprocessing is the process of producing understandable data by transforming the collected raw data [35],[36]. It enables the result of the research to be accurate and efficient by improving the quality of data [35],[36]. So, incident types and features that are relevant for this study should be selected from the raw data collected from the IMARS incident log, through interview and data analysis. Domain experts are also involved in data preprocessing on the top ten incident types and feature selection.

To do so, the following tasks are done during the data preprocessing and feature selection process. As depicted in the fig 4.2 below, the data preprocessing process includes data reduction, data cleaning, data integration, and data transformation. In the next portion, each task of the process is discussed in detail.

Figure 4.2: Data Preprocessing Process [36]

4.2.1 *Data Reduction*

As the name implies, data reduction is used to minimize the volume of data by maintaining the integrity of the original data [35],[36]. Dimensionality reduction, numerosity reduction, and data compression are some of the techniques used for data reduction [35],[36].

For this study, the dimensionality reduction technique is used to reduce the amount of raw data. Dimensionality indicates the number of attributes or features in the dataset. So in this technique, the number of attributes or features in the dataset is reduced to decrease the dimension of the dataset. Dimensionality reduction by itself has two sections: feature extraction and feature selection. In the feature extraction part, a high volume of data is reduced into a lower volume and produces a dataset with a fewer number of features. On the other hand, the feature selection segment is a process of finding a small subset that represents the original dataset [35],[36].

4.2.1.1 *Feature Extraction*

Hence, in the 6-month incident log, the amount of data extracted from IMARS contains 4531 incidents with 17 previously identified incident types. Processing such an amount of data does not cost and time efficient. In order to smooth the research work and reduce the amount of dataset, a dimensionality reduction technique is applied. In the feature extraction phase of dimensionality reduction, the top 10 incident types are selected by taking their frequency of occurrence as a major criterion. Applying this technique step enables to reduce the size of the dataset from 4531 incidents into a total of 1924 incidents. The extracted top 10 incident types are summarized in table 4.4 below.

No	Incident Type	Frequency
1	Unable open/work applications (Excel problem, Word Problem, Ppt problem, Project problem, Access problem, power geez, micro soft projector)	442
2	Unidentified network due to "IP conflict/different system cannot log on the same IP"	292
3	IPCC unable to login due to long time loading stack problem	291
4	Outlook problem unable to send and receive mail	207
5	IPCC unable to display customer information	194
6	IPCC unable to login due to "RES unknown error" " failure of open eye registration"	194
7	IPCC exit by itself	93
8	Unable to open outlook load for long time	90
9	Outlook update delay	73
10	Unable to open any file data due to corrupted /no Â compatible software	48

Table 4.4: Selected Top 10 Incident Types

4.2.1.2 *Feature Selection*

As stated in the data collection section, causes or indicators of the selected top 10 incident types and their recommended expert solution are collected from domain experts through interview and through document analysis. The quality of the data collected through interview and document analysis also needs to be maintained.

From feature selection strategies, filter strategy is selected to be used for this research work. Filter strategy select features using information gain, also known as mutual information, by measuring dependency of the two variables. In this study, filter strategy is applied on the dataset that contains the 30 causes or indicators of incident types as independent variables and the top 10 incident types as a dependent variable. Then the 30 incident causes or indicators of incident types are selected accordingly and preserved as features to model the KBS.

4.2.2 *Data Cleaning*

Data cleaning is the process of cleaning a dataset to provide accurate and complete datasets to the models [35]. This can be done by replacing missing values and by removing incorrect, incomplete, and inaccurate data [35], [36]. Handling missed and noisy or irrelevant data are the techniques that can be used in this process. Missed values and noisy data can be handled by replacing them with the probable value or by removing them totally from the dataset [35], [36].

For this study, this process runs over the reduced dataset which contains 1924 total incidents. There were missed values that cannot be replaced by the probable values, accordingly, those values are removed from the dataset and a total of 1786 incidents are preserved as a complete and accurate dataset.

4.2.3 *Data Integration*

In the case when the data is gathered from different sources, data integration is a critical step of data preprocessing. It is a technique used to consolidate the data collected from different sources into one. [35], [36] .

In this stage the data set, which is previously obtained from IMARS incident log, through interview and from data analysis preprocessed by passing through data reduction and data cleaning phases, integrated into one completed data set.

The final data set prepared after data integration is presented in the table 4.5. The incident causes or indicator values presented with zero and ones.The incident causes with the value of one indicates the presence or existence of that incident causes whereas the incident causes with the value of zero specifies the absence of the incident causes.

Table 4.5: – continued from previous page

Class Label (Top 10 Incident Types)	Recommended Solution
Unidentified network due to "IP conflict/different system cannot log on the same IP".	• Find unused IP Address •Assign unique IP address for each PC •If possible change to dynamic IP
Unidentified network due to "IP conflict/different system cannot log on the same IP".	•Contact ISD to adjust Static IP range •Assign unique IP address for each PC
Unidentified network due to "IP conflict/different system cannot log on the same IP".	•Check and reconfigure DHCP Server • Refresh Switch
IPCC unable to login due to long time loading stack problem	Add the link into compatible mode in the IE setting
IPCC unable to login due to long time loading stack problem	Restart VM
Outlook problem unable to send and receive mail	Check and adjust network connection
Outlook problem unable to send and receive mail	•Archive old mails •Delete unnecessary mails
Outlook problem unable to send and receive mail	•Recreate user profile or • Open with safe mode
Outlook problem unable to send and receive mail	• Set outlook is to be online mode •Update Outlook
IPCC unable to display customer information.	Escalate the problem to ISD and get solution
IPCC unable to display customer information.	Adjust Network Connection
IPCC unable to display customer information.	Reset IE setting
IPCC unable to display customer information.	Apply forced GP update
IPCC unable to login due to "RES unknown error failure of open eye registration"	•Reset browser •Push the Open eye license •Restart VM •Change agent open eye account •Change licensed TC
Unable to open MS applications	Activate the MS office
Unable to open MS applications	Install licensed office
Unable to open MS applications	•Recover window to previous state • Install New MS Office
Unable to open MS applications	Update OS
Unable to open MS applications	Adjust Mcafee antivirus
IPCC exit by itself	Adjust network Connection
IPCC exit by itself	•Adjust IE setting or •Add the link into compatible mode in the IE setting
IPCC exit by itself	• Adjust IE setting or • Add the link into compatible mode in the IE setting
IPCC exit by itself	Restart PC or TC
IPCC exit by itself	• Reset IE setting • Restart PC or TC
Unable to open outlook load for long time	•Recover window to previous state • Reinstall New MS Office
Unable to open outlook load for long time	Activate the MS office
Unable to open outlook load for long time	Scan the PC
Unable to open outlook load for long time	•Recreate user profile • Open with safe mode
Unable to open outlook load for long time	Adjust network Connection
Unable to open outlook load for long time	Restart PC or TC
Outlook update delay	Adjust the attachment according to the limit
Outlook update delay	Adjust network Connection
Outlook update delay	Forcibly perform GP update
Outlook update delay	Check outlook is not on offline mode and adjust the mode to be online
Unable to open any file data due to corrupted /no Â compatible software	•Scan the PC • Install licensed antivirus App • Recover the window onto previous working state
Unable to open any file data due to corrupted /no Â compatible software	• Reinstall application software • Recover the window onto previous working state
Unable to open any file data due to corrupted /no Â compatible software	Install compatible App

Table 4.5: – continued from previous page

Multiple browser open at time	IE crushed or stack	Outlook problem	Microsoft Office Licensee	PC performance problem	Outlook doesn't closed properly	Attachment file size is beyond the limit	File corrupted/attacked by virus	Application Software Corrupted	Application Software is not compatible for the file
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1

Table 4.5: – continued from previous page

Cookies and temp file	PC miss policy update	Open eye does not get a license	MS Office activation problem	Unlicenced MS Office	Corrupted MS Office application	OS Problem	Mcafee antivirus issue	System stack	IPCC system become busy
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Table 4.5: Integrated Dataset

Static IP assignment Problem	DHCP Problem	Shortage of Static IP	Compatibility issues on browser	Network (Intranet) problem	Mail size exceed	User account Problem	Outlook set to offline	IPCC - CRM Integration failure	Network interruption between TC and server side
1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

4.2.4 *Data Transformation*

Data transformation is the process of making a change in the format or structure of the data set to make it suitable for the selected model [35], [36].

In this process, the data set is transformed and coded into a format, which is suitable to build the rule-based, case-based, and ANN based knowledge base system development approach models. The detail of the data transformation process for each KBS development approach is stated in the Knowledge representation part of the Knowledge-based system model building section.

4.3 BUILDING KNOWLEDGE BASE SYSTEM MODEL

After completing the data preprocessing and feature selection step, the next phase is building the KBS model using the selected KBS development approaches. The selected three KBS development approaches are rule-based, case-based and ANN based KBS development approaches. The final dataset contains 30 causes or indicators of incident types taken as independent variables and features, the top ten incident types taken as dependent variables or class labels, and a total of 1786 incidents taken as instances.

In order to build the KBS model there is a general process aligned with KBS architecture as stated in chapter two that needs to be accomplished.

Each step taken to build the KBS model using the selected KBS development approaches is explained in the coming subsection.

4.3.1 *Rule-Based knowledge-based system model building*

Knowledge Acquisition

The steps mentioned in the chapter 2 knowledge acquisition process are applied for each KBS development approach model building to extract the relevant knowledge.

As mentioned in the data collection section, knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the rule-based KBS model.

knowledge Representation

For the purpose of this study, the preprocessed domain knowledge is changed into the form of "IF-THEN "rules. In this case, the "IF" side holds class features, which are the possible causes or indicators of the incident types, and the "THEN" side of the statement states the class label which is the identified incident type that might occur when the listed causes or indicators exist with the appropriate expert solutions for it.

From the domain knowledge, 32 rules are derived to identify the class labels, which are the selected incident types. In addition, the rules are designed based on the class features, which are identified causes or indicators of each incident type with the expert's solution. Sample rules out of 32 rules, which are included in the rule-based reasoning KBS model are mentioned below.

Rule:
IF Cookies and temp files exist,
  THEN IPCC unable to display customer information incident occur,
  SOLUTION: Reset IE setting
    IF PC miss policy update,
    THEN IPCC unable to display customer information incident occur,
    SOLUTION: Apply forced GP update
      IF IPCC - CRM Integration fail,
      THEN IPCC unable to display customer information incident occur,
      SOLUTION: Escalate the problem to CRM Team to recover
        the integration

Rule:

IF there is a Static IP assignment problem,

THEN Unidentified network due to "IP conflict/different system cannot log on the same IP" incident occur,

SOLUTION: Find unused IP Address and Assign unique IP address for each PC or If possible change to dynamic IP

IF there is shortage of static IP,

THEN Unidentified network due to "IP conflict/different systems cannot log on the same IP" incident occur,

SOLUTION: Contact ISD to adjust Static IP range and assign unique IP address for each PC

IF DHCP problem exist,

THEN Unidentified network due to "IP conflict/different system cannot log on the same IP" incident occur,

SOLUTION: Check and reconfigure DHCP Server
Refresh Switch

Building Rule-Base knowledge based system Model

In this case, the knowledge-based system model is built using the rule-based KBS development approach. The rule-based knowledge-based system model is built using the python programming language. The rule engine is designed as a KB based on the 32 rules derived in the knowledge representation step. The system is also designed and implemented to infer the value of the input features as causes and deduce the appropriate class labels, which are incident types, and recommended expert solutions by cross-checking the input regardless of the rules.

4.3.2 *Case-Based knowledge based system model building*

Knowledge Acquisition

The steps mentioned in the chapter 2 knowledge acquisition process are applied for each KBS development approach model building to extract the relevant knowledge.

As mentioned in the data collection section, knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the case-based KBS model.

knowledge Representation

In this study, in order to build the case-based KBS model, the preprocessed domain knowledge is converted into cases or experiences. From the domain knowledge, 32 cases are drawn based on the 30 features obtained from incident causes or indicators, to identify the most similar class labels, which are the incident type with the appropriate expert solution.

Sample cases, out of 32 cases that are included in the case-based KBS model, are mentioned below.

Case 1:
Static IP assignment problem exist , there is a DHCP Problem, there is no shortage of Static IP, there is no compatibility issues on browser, the intranet and the network working well , mail size is not exceed to the limit, user account has no problem, outlook is not set to offline, IPCC and CRM Integration is working well, there is no network interruption between TC and server side , cookies and temp files are cleaned properly, PC properly update its policy and has no performance problem , open eye get its license, MS Office activate on time accurately, the installed MS office is licensed, MS Office application is not corrupted , there is no problem with the OS, Mcafee antivirus has no issue , IPCC system is working well without stack and busy , multiple browser doesn't open at time, IE is working well, there is no outlook problem and outlook is closed properly , mail attachment file size is not beyond the limit ,the attached file is not attacked by virus, application Software is compatible with the attached file and not corrupted.

The incident type is: Unidentified network due to IP conflict
different system cannot log on the same IP

The solution is: Find unused IP Address and
assign unique IP address for each PC or
if possible change the static IP to dynamic

Case 21:

There is no static IP assignment problem , there is no a DHCP Problem, there is no shortage of Static IP, there is no compatibility issues on browser, the intranet and the network working well , mail size is not exceed to the limit, user account has no problem, outlook is not set to offline, IPCC and CRM Integration is working well, there is no network interruption between TC and server side , cookies and temp files are cleaned properly, PC properly update its policy and has no performance problem , open eye get its license, MS Office activate on time accurately, the installed MS office is licensed, MS Office application is not corrupted , there is no problem with the OS, Mcafee antivirus has no issue , IPCC system is became stack and busy , multiple browser doesn't open at time, IE is working well, there is no outlook problem and outlook is closed properly , mail attachment file size is not beyond the limit ,the attached file is not attacked by virus, application Software is compatible with the attached file and not corrupted.

The incident type is: IPCC exit by itself

The solution is: Adjust IE setting or
Add the link into compatible mode in the IE setting or
Restart PC or TC

Building Case-Base knowledge based system Model

In this case, the knowledge-based system model is built using the case-based KBS development approach. The case-based knowledge-based system model is built using the python programming language and based on the case-based cycle process. The case-based library is designed as a KB based on the 32 previous cases prepared in the knowledge representation step. The system is also designed to prepare the new case using the value of the input features and infer to deduce class labels, which are incident types with recommended expert solutions, by cross-checking the new case with existing cases and suggesting the one with the highest similarity.

4.3.3 *ANN-Based knowledge based system model building*

Knowledge Acquisition

The steps mentioned in the chapter 2 knowledge acquisition process are applied for each KBS development approach model building to extract the relevant knowledge.

As mentioned in the data collection section, knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the ANN-based KBS model.

knowledge Representation

In this study, 30 features, which are causes and indicators of incident types, are used as an input feature and the top ten incident types with their expert solution are used as a class label to build ANN-based KBS model.

In order to represent the knowledge for the ANN based KBS model the preprocessed knowledge, which contains the 30 input features and the 10 class labels, are encoded from text variables into dummy variables. Dummy variables are variables that have zero and one values. The variable with the value of one indicates the presence or existence of that variable whereas the variable with the value of zero specifies the absence of the variable. Then the knowledge is coded accordingly to predict the ten class labels.

Knowledge represented for the ANN based KBS model is the same as table 4.5.

Building ANN-Based knowledge based system Model

From the available ANN types, a feed-forward neural network that is suitable for classification is selected for this study. In this case, the knowledge-based system model is built using the ANN-based KBS development approach. The ANN-based knowledge-based system model is built using the python programming language.

The model is built using the selected 30 features which are the causes or indicators of the incident types. So, the input layer has 30 neurons to handle the 30 features.

The number of the epoch, hidden layers, and neurons across each hidden layer affect the performance of ANN-based KBS model. So ANN-based approach is built and trained using different numbers of an epoch, hidden layers, and neurons. An epoch is a number of cycles used to train the neural network with all the training datasets. Then finally, the number of an epoch, hidden layers, and neurons that came up with higher performance is taken for this study. From the model-building progress, the below points are observed and a decision is made accordingly.

- having two and above hidden layers produce the same performance result
- even if there is no standard formula to decide on the number of neurons that can be available on each layer, there are basic concepts to be considered on selecting the hidden neurons.
  - Number of neurons in each hidden layer should be between the size of the input layer neurons and the size of the output layer neurons.
  - Number of neurons in each hidden layer should be less than twice the size of input layer neurons
  - having a number of neurons that satisfies the above concepts per each hidden layer makes the model in a better performance
- using 70% of the dataset as a training dataset produces a better performance
- having 100 and above epoch produce the same performance result

Therefore as a conclusion

- Two hidden layers are used to process the input features using relu as an activation function.
- 25 and 15 hidden neurons are used for the first and second hidden layers
- Out of the 1786 instances 70% of the data set is used to train the ANN based KBS model
- Out of the 1786 instances 30% of the data set is used to test the ANN based KBS model

- 200 epoch cycles are used to train the ANN based KBS model

Related kinds of literature also support this conclusion [20].

Hence the output of the ANN based KBS model consists of 10 class labels, which are the top 10 incident types with their appropriate experts' solution, the output layer is designed to have 10 neurons and process the output using softmax as an activation function.

4.4 PERFORMANCE EVALUATION METRICS

After building the selected knowledge-based system development approaches model, the next step is evaluating and comparing the performance of the approaches model towards identifying the class labels, which are incident types with recommended expert solutions.

A number of standard metrics are available to measure and compare the performance of knowledge based system development approaches [37], [38], [39], [40]. Among them, six performance evaluation metrics are selected and used for this study to evaluate and compare the performance of the KBS development approaches. These are execution time, accuracy, F-Measure, precision, recall, error rate and mean square error (MSE). The evaluation is done by comparing the result of the KBS development approaches against the domain expert's decision on the class label identification or incident type identification with solution recommendation.

The size of the test dataset applied to evaluate the models contains a total of 1786 instances, incidents.

4.4.1 *Execution Time*

Execution time is used as a performance evaluation metric, which measures the time taken by the system to respond to the user request. It is used to measure the response time of KBS development approaches.
The time required to build and execute the rule-based, case-based, and ac{ANN-

based KBS development approach models is measured in seconds. And the models' response time taken to predict the class labels, which are the top 10 incident types, after accepting the input features is compared.

4.4.2 *Accuracy*

Accuracy is the simplest performance evaluation metric. It is defined as the ratio of correctly identified or predicted instances with respect to the total number of the dataset, multiplied by 100 [37], [39]. The mathematical representation of accuracy is shown in equation 4.1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{4.1}$$

Where;

TP: Stands for true positive and indicates the number of instances in which the algorithm correctly predicted as positive or yes and the actual label is also positive or yes.

TN: Stands for true negative and indicates the number of instances in which the algorithm correctly predicted as negative or no and the actual label is also negative or no.

FN: Stands for false-negative and indicates the number of instances in which the algorithm incorrectly predicted to be negative or no and the actual label is positive or yes.

FP: Stands for false-positive and indicates the number of instances in which the algorithm incorrectly predicted to be positive or yes and the actual label is negative or no.

The rule-based, case-based, and ANN-based KBS development approach models' ability on how correctly predict the exact class label, the top 10 incident types, for the 1786 instances. Based on the formula the accuracy of the models is measured by computing TN, TP, FN, and FP. In this study, there are 10 class labels named as class label A to class label J. Let the class label of the instance is class label A, so according to this study TN, TP, FN, and FP means:

- True positive {TP}, the instance class label is predicted by the model as class A and the actual class label is also Class A
- True Negative {TN}, the instance class label is predicted by the model as class label B or C or D or E or F or G or H or I or J class label, and the actual class label is B or C or D or E or F or G or H or I or J
- False-negative {FN}, the instance class label is predicted by the model as class label B or C or D or E or F or G or H or I or J class label, and the actual class label is A
- False-positive {FP}, the instance class label is predicted by the model as class label A and the actual class label is B or C or D or E or F or G or H or I or J

F-measure, precision, recall and error rate performance of the models were also evaluated using the above definition of TP, TN, FP, and FN.

4.4.3 *F-Measure*

F-measure is one of the commonly used performance evaluation metrics. It is the harmonic mean between precision and recall [37], [40]. It shows the effectiveness of the algorithm on how correctly predict instances without missing an important number of instance[40]. Equation 4.2 tells how F-measure is calculated.

$$\begin{aligned} F-measure &= \frac{2 * Precision * Recall}{Precision + Recall} \\ &= \frac{2TP}{2TP + FP + FN} \end{aligned} \tag{4.2}$$

Precision: measures the ratio of correctly predicted instances as positive with respect to the sum of correctly predicted instances as positive and incorrectly predicted instances as positive [37], [40]. If the precision is high, it indicates that the instance miss-predicted by the model is less in number (small FP). It can be calculated using equation 4.3.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.3}$$

Recall: measures the ratio of correctly predicted instances as positive with respect to the number of all relevant samples (the sum of correctly predicted instances as positive and incorrectly predicted instances to be Negative) [37], [40]. It can be calculated using equation 4.4.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.4}$$

4.4.4 *Error Rate*

The error rate is an error measure that can be obtained by subtracting the algorithm accuracy performance rate from one(100%). In another form, it can be defined as a ratio of an incorrectly evaluated instance over the total number of evaluated instances[37]. The error rate can be calculated using equation 4.5.

$$\text{ErrorRate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.5}$$

4.4.5 *Mean Square Error (MSE)*

Mean square error is one of the error rate measurement which measures the error in a mean squared form between the predicted result and the actual result

[37], [41]. To have a good performance, the algorithm must score small MSE. The equation to obtain MSE is defined as equation 4.6 [37]:

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(p_j - A_j)^2 \tag{4.6}$$

Where;

Pj: is the predicted value obtained from the algorithm for instance j,

Aj: is actual target value for instance j

n: is the total number of instances

The rule-based, case-based, and ANN-based KBS development approach models' errors between the predicted class label and the actual class label. Based on the formula MSE of the models is measured by computing the parameters Pj, Aj and n. According to this study Pj, Aj and n means:

- Pj: the class label of instance j predicted by the model
- Aj: the actual class label of instance j
- n: the total number of instances, which is 1786

5

RESULTS AND DISCUSSION

Knowledge base systems have different performance. The development approaches used in KBS inference engine for reasoning the conclusion and to construct the knowledge base engine affect the performance of KBS. This study shows how the performance of the KBS depends on the applied KBS development approach. The study compares the performance of KBS model build using rule, case, and ANN-based approaches. Performance evaluation metrics such as execution time, accuracy, F-measure, precision, recall, error rate, and Mean Square Error (MSE) are used to compare the KBS development approaches.

The same dataset is used to test all KBS development approach models. The data set contains 30 causes or indicators of incident types taken as independent variables and features, the top ten incident types taken as dependent variables or class labels, and a total of 1786 incidents taken as instances.

The result of the study is analyzed based on the outcome of the execution time taken by the system to respond to the input dataset, the accuracy, error in class label or incident type identification using error rate, MSE, and F-measure including precision and recall.

Execution Time
As stated in table 5.1 and fig 5.1 unlike rule-based and case-based, ANN based KBS development approach takes the lowest time, which is 0.95 seconds to build and execute the model to predict or identify the class labels or incident types with recommended expert solution. Rule-based and case-based KBS development approaches execute in 2.7 and 6.8 seconds respectively. Compared to case-based approach, rule-based approach execute with slightly minimum time.

KBS Development approaches	Execution Time (*in second*)
Rule Based Reasoning	2.7
Case Based Reasoning	6.8
ANN Based Reasoning	0.95

Table 5.1: Execution time taken by each KBS development approaches

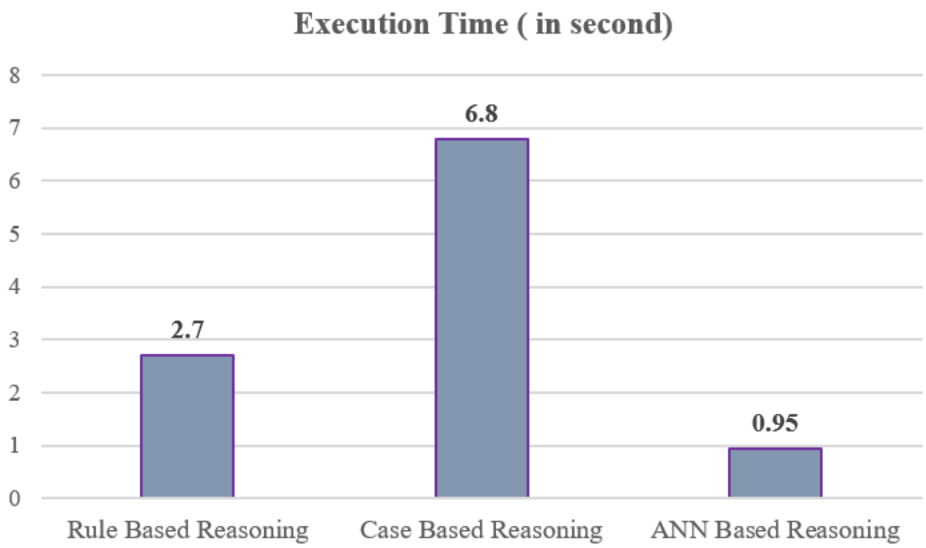


Figure 5.1: Execution time metrics in second

Accuracy

An accuracy of 98.1 % achieved by the rule-based KBS development approach. Whereas the case-based and the ANN based development approaches record accuracy of 96.8 % and 97.45 % respectively. In this result, the case-based development approach is the least performer. Compared to others, the rule-based KBS development approach has better accuracy in predicting or identifying the accurate class label or incident type with recommended appropriate solutions. The result of the accuracy performance measurement is presented below in table 5.2 and fig 5.2.

KBS Development approaches	Accuracy (%)
Rule Based Reasoning	98.6
Case Based Reasoning	96.8
ANN Based Reasoning	97.45

Table 5.2: Accuracy performance of each KBS development approaches

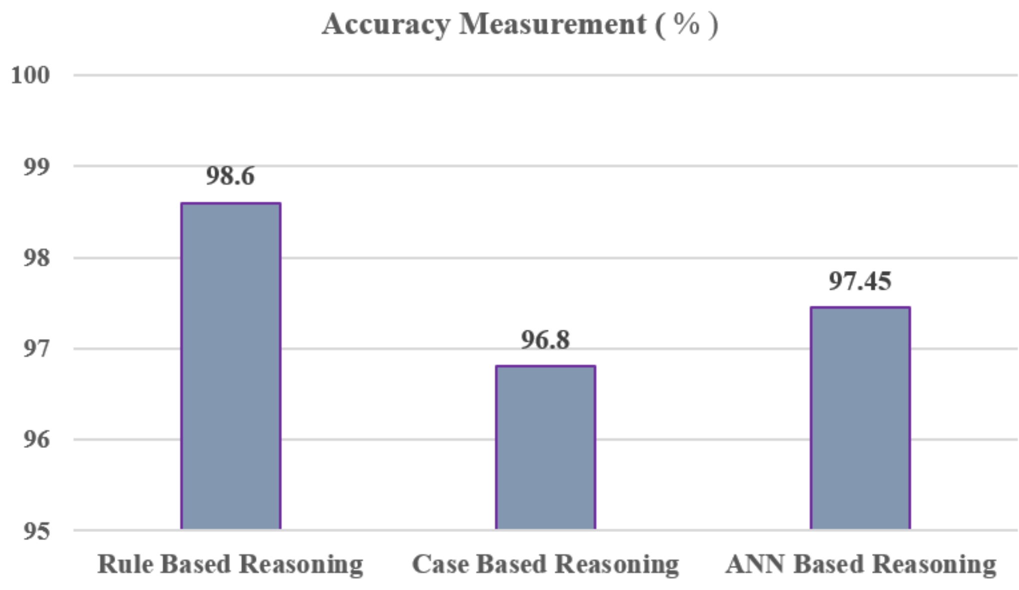


Figure 5.2: Accuracy performance of each KBS development approaches

The F-measure, precision, and recall presented in table 5.3 and figure 5.3 below shows the same result as the overall accuracy can show in table 5.2 and figure 5.2. The result clearly shows that the rule-based development approach has the highest F-Measure, precision, recall. This indicates that the rule-based development approach is more efficient to correctly predict or identify the class labels or incident types and respective recommended expert solutions with less probability of miss predicting or identifying the class labels.

KBS Development approaches	F-Measure	Precision	Recall
Rule Based Reasoning	0.99	0.99	0.99
Case Based Reasoning	0.97	0.98	0.97
ANN Based Reasoning	0.97	0.98	0.97

Table 5.3: F-measure, Precision and recall performances

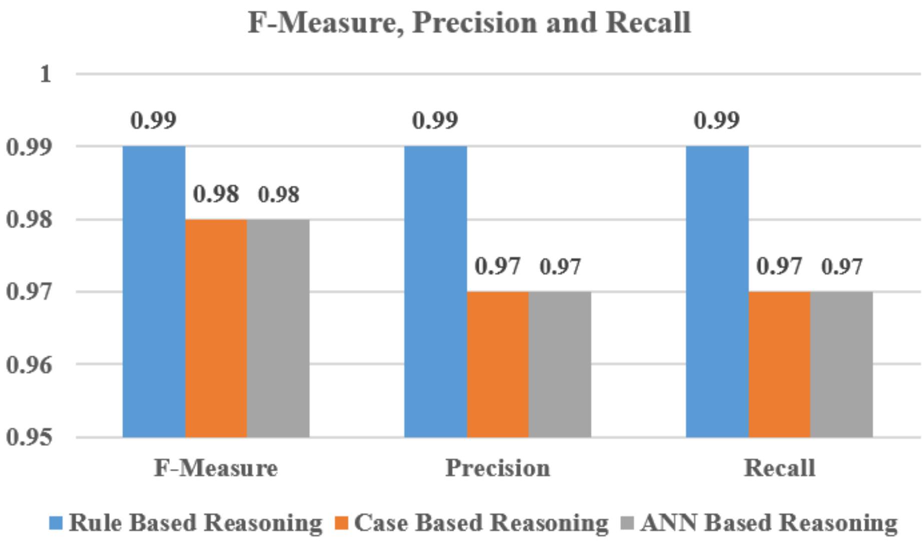


Figure 5.3: F-measure, Precision and recall performances

Errors

The rule-based KBS development approach is the best approach in minimizing errors, hence it scores the minimum error rate and MSE which is 1.4 % and 0.187. Next to the rule-based approach, ANN based development approach has the minimum error rate and MSE which is 2.54 % and 0.35. Whereas the case-based development approach produces major errors in predicting or identifying the class labels or incident type with recommended expert solution by recording a 3.19 % error rate and 0.49 MSE. The result of the error performance measurements is presented below in table 5.4 and fig 5.4.

KBS Development approaches	Error rate	Mean Square Error (MSE)
Rule Based Reasoning	1.4	0.187
Case Based Reasoning	3.19	0.49
ANN Based Reasoning	2.54	0.35

Table 5.4: Errors in KBS development approaches

Figure 5.4: Errors in KBS development approaches

In conclusion, after observing the overall performance results. The rule-based approach is the best KBS development approach in accurately predicting or identifying the incident types and recommending the appropriate solutions with minimum error. In another way, the ANN based approach is the best KBS development approach to build a KBS which has a fast response time. The conclusion is driven according to the evaluation metrics used for this thesis.

6 CONCLUSION AND RECOMMENDATION

6.1 conclusion

Telecom companies, including ethio telecom, challenged to provide quality and consistent operation support and maintenance without timely availability of domain experts. This can be efficiently be managed using the KBS. KBSs are designed to solve a problem using experts' decision-making procedures and abilities by simulating their knowledge, methods, and experience. There are several development approaches used to build KBS. To select and use one of them, companies or individuals need to make sure that the selected approach has the best performance.

Researchers try to evaluate the performance of different KBS development approaches to investigate the best approach. However, the performance of such development approaches is not studied in the context of telecom companies, especially ethio telecom. In addition, most of them only evaluate the accuracy of the approaches even if there are additional metrics that have an impact on the incident handling time of IT technicians and must be seen to investigate fast and accurate KBS development approaches.

This study is conducted to identify the best KBS development approaches appropriate for ethio telecom, which has the best performance in accuracy, execution time, F-measure, precision, recall and produce less error in error rate and MSE. Rule-based, case-based, and ANN based KBS development approaches are selected and compared on their ability in predicting or identifying class labels or incident types and propose an expert solution.

The ANN-based KBS development approach model achieves a better performance in execution time evaluation metrics by building and executing the model and presenting the result in 0.95 seconds. Rule-based and case-based KBS development approaches models accomplish 2.7 and 6.8 seconds in the execution time respectively. In another way, the rule-based KBS development approach model scores the best in accuracy measurement, which is 98.1 %. In these metrics, the case-based and ANN based development approach models record 96.8 % and 97.45 %. In this result, the case-based development approach is the least performer. Compared to others, the rule-based KBS development approach has better accuracy in predicting or identifying the class labels or the accurate incident type with recommended appropriate solutions. As the same as the overall accuracy, the rule-based KBS development approach model has the highest F-Measure, precision, recall performance. The rule-based KBS development approach model also achieves the best performance in minimizing errors, hence it scores the minimum error rate and MSE which is 1.4 % and 0.187. ANN based development approach model produces 2.54 % error rate and 0.35 MSE. Whereas the case-based development approach model produces a 3.19 % error rate and 0.49 MSE.

According to the result of the performance evaluation, the rule-based approach is the best KBS development approach in accuracy, error rate as well as in MSE evaluation metrics. AS well, ANN based KBS development approach is the best KBS development approach to build a fast response time KBS.

In the ethio telecom context, the response time of the KBS is related to the handling time of incidents and gives a high priority. Accuracy of the KBS to suggest the exact class labels or incident type with related expert solutions, also affect the handling time of an incident. If a high priority is given to response time, ANN based development approach is a better solution to build a KBS for ethio telecom. While a high priority is given to accuracy and error metrics, the rule-based development approach is the better one.

It is known that all evaluation metrics have an impact on the handling time of an incident and are taken as mandatory metrics. So, as a conclusion, this study proposes a hybrid KBS development approach by combining both rule-based and ANN based KBS development approach to be benefited from both approaches. The proposed approach enables ethio telecom to have a fast and accurate KBS with minor errors in identifying the class label or occurred incident types with recommended expert solutions. Indirectly, the proposed approach allows ethio telecom to resolve its challenge occurred by the time its domain expert can't exist and enable to provide consistent operation support with available IT technicians by adopting expert solutions.

6.2 recommendations for future work

This study suggests that ethio telecom must build KBS by combining rule-based and ANN based KBS development approaches to have a fast and accurate KBS. But there may be a lot of ways on how to build a hybrid KBS by combining the rule and ANN based approaches. So studying the better combination approaches can be taken as future work.

REFERENCES

[1] E. F. Vail, "Knowledge mapping: Getting started with knowledge management," *Inf. Syst. Manag.*, vol. 16, no. 4, pp. 1–8, 1999.

[2] B. J. Bowman, "Building knowledge management systems," in *IS Management Handbook*, Auerbach Publications, 2013, pp. 877–892.

[3] B. Grasic and V. Podgorelec, "Developing knowledge management systems: Approaches, technologies and methods," in *Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications*, 2011, pp. 207–212.

[4] E. M. Khaemba, "Factors affecting knowledge management implementation in telecommunication industry: A case study of safaricom limited," *International Journal of Engineering Research & Technology*, vol. 2, no. 6, pp. 1358–1366, 2013.

[5] X. Huang and J. R. Jensen, "A machine-learning approach to automated knowledge-base building for remote sensing image analysis with gis data," *Photogrammetric engineering and remote sensing*, vol. 63, no. 10, pp. 1185–1193, 1997.

[6] L. Atymtayeva, K. Kozhakhmet, and G. Bortsova, "Building a knowledge base for expert system in information security," in *Soft computing in artificial intelligence*, Springer, 2014, pp. 57–76.

[7] E De Kock, "Chapter 6–expert systems and knowledge acquisition," *University of Pretoria etd*, pp. 107–125, 2003.

[8] K. Tripathi, "A review on knowledge-based expert system: Concept and architecture," *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications*, vol. 4, pp. 19–23, 2011.

[9] M. D. C. Tongco, "Purposive sampling as a tool for informant selection," *Ethnobotany Research and applications*, vol. 5, pp. 147–158, 2007.

[10] H. Alshenqeeti, "Interviewing as a data collection method: A critical review," *English linguistics research*, vol. 3, no. 1, pp. 39–45, 2014.

[11] R. Burnashev, I. Amer, and A. Enikeev, "Expert system building tools based on dynamically updated knowledge," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1352, 2019, p. 012 008.

[12] A. I. Enikeev, R. A. Burnashev, and G. Z. Vakhitov, "Software tools and techniques for the expert systems building," in *Fourth International Congress on Information and Communication Technology*, Springer, 2020, pp. 191–199.

[13] E. Adetiba, T. John, A. Akinrinmade, F. Moninuola, O. Akintade, and J. Badejo, "Evolution of artificial intelligence languages, a systematic literature review," *arXiv preprint arXiv:2101.11501*, 2021.

[14] G. Neumann, "Programming languages in artificial intelligence," *Encyclopedia of Information Systems*, pp. 31–45, 2002.

[15] J. Sundnes, *Introduction to Scientific Programming with Python*. Springer Nature, 2020.

[16] B. Heinold, "A practical introduction to python programming," *Creative Commons Attribution*, 2012.

[17] L. B. Kaimal, A. R. Metkar, and G Rakesh, "Self learning real time expert system," *International Journal on Soft Computing, Artificial Intelligence and Application (IJSCAI)*, vol. 3, no. 2, pp. 13–25, 2014.

[18] S. Mandal, S. Chatterjee, and B. Neogi, "Diagnosis and troubleshooting of computer faults based on expert system and artificial intelligence," *International Journal of Pure and Applied Mathematics*, vol. 83, no. 5, pp. 717–729, 2013.

[19] M. M. Mustaffa, A. Shibghatullah, A. Basari, and B Hussin, "Diagnosing computer hardware failures using expert system (rule-based technique)," in *International Symposium on Research in Innovation and Sustainability*, 2014.

[20] Y. Wang, Q. Li, M. Chang, H. Chen, and G. Zang, "Research on fault diagnosis expert system based on the neural network and the fault tree technology," *Procedia Engineering*, vol. 31, pp. 1206–1210, 2012.

[21] S. Bringsjord and B. Schimanski, “What is artificial intelligence? psychometric ai as an answer,” in *IJCAI*, Citeseer, 2003, pp. 887–893.

[22] J. E. Aronson, T.-P. Liang, and R. V. MacCarthy, *Decision support systems and intelligent systems*. Pearson Prentice-Hall Upper Saddle River, NJ, USA: 2005, vol. 4.

[23] G. Yost and A. Newell, “A problem space approach to expert system specification.,” in *IJCAI*, Citeseer, vol. 89, 1989, pp. 621–627.

[24] J. Prentzas and I. Hatzilygeroudis, “Categorizing approaches combining rule-based and case-based reasoning,” *Expert Systems*, vol. 24, no. 2, pp. 97–122, 2007.

[25] B. Buchanan and R. Duda, *Principles of rule-based expert systems. stanfrod, usa*, 1982.

[26] A. Aamodt and E. Plaza, “Case-based reasoning: Foundational issues, methodological variations, and system approaches,” *AI communications*, vol. 7, no. 1, pp. 39–59, 1994.

[27] A.-B. M. Salem, “Case based reasoning technology for medical diagnosis,” *World academy of science, engineering and technology*, vol. 31, no. 2007, pp. 9–13, 2007.

[28] S. M. Saif, M. Sarikhani, and F. Ebrahimi, “An expert system with neural network and decision tree for predicting audit opinions,” *IAES International Journal of Artificial Intelligence*, vol. 2, no. 4, p. 151, 2013.

[29] D. Arditi and O. B. Tokdemir, “Comparison of case-based reasoning and artificial neural networks,” *Journal of computing in civil engineering*, vol. 13, no. 3, pp. 162–169, 1999.

[30] G. Liu, X. Yang, Y. Ge, and Y. Miao, “An artificial neural network-based expert system for fruit tree disease and insect pest diagnosis,” in *2006 IEEE International Conference on Networking, Sensing and Control*, IEEE, 2006, pp. 1076–1079.

[31] G. L. Team, *Types of neural networks and definition of neural network*, https://www.mygreatlearning.com/blog/types-of-neural-networks/, 2021.

[32] K. Maladkar, *6 types of artificial neural networks currently being used in machine learning*, https://analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/, 2018.

[33] K. Goyal, *The 7 types of artificial neural networks ml engineers need to know*, https://www.upgrad.com/blog/types-artificial-neural-networks-in-machine-language/, 2019.

[34] I. A. Basheer and M. Hajmeer, "Artificial neural networks: Fundamentals, computing, design, and application," *Journal of microbiological methods*, vol. 43, no. 1, pp. 3–31, 2000.

[35] S. Anunaya. "Data preprocessing in data mining -a hands on guide." (2021).

[36] A. Joby. "What is data preprocessing? 4 crucial steps to do it right." (2021).

[37] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.

[38] M. Steurer, R. J. Hill, and N. Pfeifer, "Metrics for evaluating the performance of machine learning based automated valuation models," *Journal of Property Research*, pp. 1–31, 2021.

[39] J. Fürnkranz and P. A. Flach, "An analysis of rule evaluation metrics," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 202–209.

[40] A. K. Das, "Introduction to research evaluation metrics and related indicators," in *Open Access for Researchers, Module 4: Research Evaluation Metrics*, UNESCO, Paris, 2015, pp. 1–18.

[41] A. Jierula, S. Wang, T.-M. Oh, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *Applied Sciences*, vol. 11, no. 5, p. 2314, 2021.

A

APPENDIX

Comparative Analysis of Knowledge Base System Development Approaches in the context of ethio telecom

Biruktait Fikre Hailu
School of Electrical and Computer Engineering
Addis Ababa Institute of Technology
Addis Ababa University
Addiss Ababa, Ethiopia
birktfkr@gmail.com

***Abstract*—For a successful business, telecom companies need to provide consistent, quality, and timely maintenance and support to their customers. This is difficult to achieve without the timely availability of domain experts that can efficiently be managed using the knowledge Base System. Several approaches are applied to implement a knowledge Base System. Among them, this paper selects rule-based,case-based, and artificial neural network-based knowledge Base System development approaches; compares their performance in the context of the operation and maintenance process of ethio telecom, and proposes the better approach.**

Related performance studies in the context of telecom companies only focused on the accuracy performance of the approaches, there is none related studies for ethio telecom. This paper includes execution time, accuracy, F-measure, precision, recall, error rate and mean square error metrics that have an impact on incident handling time as performance evaluation metrics. Python programming language is used to model, evaluate and compare the selected approaches.

The performance results show that the Artificial Neural Network-based knowledge Base System development approach has a better performance in terms of building and executing the model, taking a minimum time of 0.95 seconds. Furthermore, rule-based approach accomplishes the highest accuracy that is 98.6 %, with minimum error in identifying the incident type. Based on the evaluation results on the response time or incident handling time, a hybrid knowledge Base System development approach from the Artificial Neural Network and rule-based knowledge Base System development approach can provide the best aggregate result.

***Index Terms*—Knowledge Base System, Rule-based development approach, Case-based development approach, Artificial Neural Network-based development approach**

I. INTRODUCTION

The term knowledge is defined as a high-value form of information, which provides a competitive significance to a company and is mostly used for decision and action support [1], [2]. It is combined with company policy, strategy, rules, procedures, expert experience, interpretation, context, reflection, and skills. There are two basic types of knowledge. Implicit or tacit knowledge resides in people as mental models, experiences, and skills. The other one is explicit knowledge, which can be communicated and captured from formal models, rules, and procedures [1], [2].

Knowledge is gradually taken as an organization's valuable resource and becomes a key to organizational success [3], [4]. To have a successive journey and drive as competitive, companies are required to manage their organization's knowledge [3], [4].

Knowledge Management(KM) is a collection of formal processes that direct on how a piece of knowledge is captured, created, distilled, organized, stored, disseminated, utilized, and retained throughout the organization [2]- [4]. Managing both, implicit and explicit knowledge of a company provide enormous benefits [1]- [4].

The KM process by itself is not technical. Information Technology (IT) emerges a computer system with an intelligent technical means to manage an organization's knowledge; known as a knowledge management system (KMS), knowledge base system, or expert system (ES) [1]- [6].

KBS and ES are used interchangeably, to express artificial intelligence (AI) based systems designed to solve a problem using an expert's decision-making procedure and ability by simulating knowledge, methods, and experience of experts [1]- [7].

The value of a KMS or KBS is recognized by different organizations and applied to problems in so many fields. From research, fields like engineering, electronics, law, agriculture, medicine, manufacturing, metrology, geology, physics, forestry, finance, IT, and telecommunication are mostly mentioned fields, which implement KMS [3] - [7].

Researchers propose different approaches like rule-based, semantic networks or ontology-based, case-based, artificial neural network (ANN) Based, frame-based, object-oriented based, and script-based approaches to implement KBS or ES [3] - [7]. Each approach has its own advantages, disadvantages, and performance. Finding a suitable approach, to build and implement the best KBS for a specific company, is not an easy task and needs deep investigation.

II. PROBLEM DESCRIPTION

Ethio telecom, including other telecom companies, implements the latest technologies to provide modern service for its customers. The company also continues to build and expand its network infrastructure, implement applications and systems that can enable the company to offer recent services and technologies in a good quality of services. For a successful business, ethio telecom needs to sustain a good quality of service by providing consistent and quality maintenance and support for their customers. This is succeeded by hiring and training IT experts. But to maintain telecom services, to handle incidents that might happen on the service and equipment, and to provide timely support for customer requests, IT experts, domain experts, need to be timely available. If IT experts can't be available due to different reasons, telecom companies encounter difficulties to provide a sustainable quality of service. Such types of cases forced the support departments to handle customer problems with the remaining staff, which may not have enough knowledge on incident handling or with newly hired ones. This makes the company to face a challenge in providing consistent operation support by handling the same incidents and customer requests with a lower capability as compared to its previous capability.

To sustain the previous way of incident handling and to obtain domain expert trends of handling, support staff will try to contact experts through different means of communication. domain experts may not be available at the required time or explain the solution explicitly; which may take extra time for understanding. This will result in handling the same problem with much time, which causes an increment in a customer complaint, decrease customer satisfaction, and avert the support and the company from meeting its target achievement. As a solution, from the previous and current trend, the support department facilitates additional training. This will cause the company to spend extra expenditure and needs additional time to prepare, schedule, and perform the training. These all the mentioned solutions are not time and cost-effective.

However, this can efficiently be managed using KBS by collecting, storing, sharing, and managing domain experts' knowledge and experience on solving problems and decision-making ability. To do so, there are several KBS development approaches and techniques, that help to manage knowledge and facilitate expert knowledge sharing, proposed by researchers. Studying the performance of the available KBS development approaches is needed to select the appropriate one. Yet there are limited KBS development approaches performance studies in the context of telecom companies and none for ethio telecom. Most of the studies are also focused only on the accuracy performance of the approach and other relevant performance metrics that have an impact on incident handling are not addressed well.

There is a need for a study that can address such research gaps and difficulties in identifying a better KBS development approach suitable for ethio telecom to solve the mentioned problem. This research focuses on proposing a better KBS development approach suitable for ethio telecom operation and maintenance process by studying and comparing KBS development approach performance that has an impact on incident handling other than accuracy.

III. Knowledge Base System

A. Introduction

The knowledge-based system also known as an expert system, is an artificial intelligence-based system designed to solve a problem using an expert's decision-making procedure and ability by simulating knowledge, methods, and experience of experts [1] - [6].

The knowledge-based system uses human expertise domain knowledge, stored in an intelligent computer program, called knowledge base, then drives expert knowledge when users need access to the knowledge. It can consult by giving a suggestion or advice like human experts on how anyone can solve a problem. KBS enables to save money, time, and effort by making several experts' knowledge to be available anywhere and anytime without the need of the experts' physical existence. It also smoothens the consistency of work by providing experts' knowledge and making anyone function in a productive manner at a higher level as an expert.

B. Architecture of Knowledge Based System

KBS as a system consists of an interconnected system component with their own roles working together to provide the required functionality. As shown in the architecture of KBS presented in fig 1, the essential components of the KBS are knowledge acquisition, knowledge representation,knowledge base (KB), inference engine, and user interface(UI) module [7], [23].

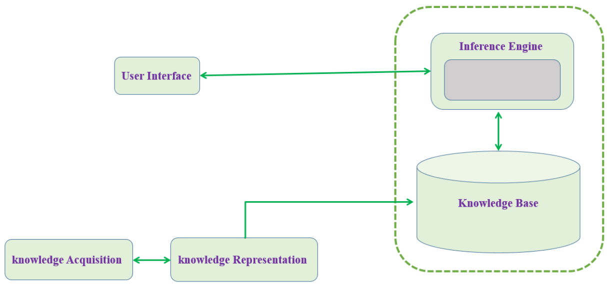


Fig. 1. Architecture of a Knowledge-Based System [23]

1) Knowledge Acquisition: Knowledge acquisition(KA) is the process of extracting relevant knowledge from specified data sources such as domain experts, documents like books, manuals, and journals [7] – [23]. KA extract knowledge to accumulate and transfer experts' knowledge and problem-solving skill into computer programs and system users. Potential data sources of knowledge include domain experts', domain-specific literature and research, textbooks, databases, and the like. Techniques such as interviews, data analysis, questioners, and observations can be used to acquire knowledge from potential data sources. KA includes eliciting and structuring of knowledge obtained from different data sources [7] – [23].

The KA process determines the performance, usefulness, and trustworthiness of KBS. So, more attention must be given to the effectiveness of the techniques used to acquire reliable, accurate, and valid knowledge [7]- [23].

In order to model the KBS efficiently, the knowledge engineer must carry out the following steps sequentially during KA process [23].

- Elicit the data obtained from domain experts and document analysis
- Interpret the acquired knowledge and conclude domain expert knowledge and reasoning process
- Construct the model by structuring the expert's knowledge
- Repeat step I-IV until the KBS became fully functional

2) Knowledge Representation: After extracting the required domain knowledge, knowledge must be encoded in a systematic means [7]- [23]. This implies a knowledge representation(KR). KR includes preparing a knowledge map and the documenting, structuring, encoding of acquired knowledge. KR is the process of transforming and encoding the acquired knowledge into a systematic means or computer understandable form to make the knowledge suitable for the selected KBS development approaches [7] – [23]. Through the process of knowledge representation knowledge engineer makes the knowledge to be suitable for the designed knowledge base. KR differs based on the selected KBS development approaches. Different KBS development approaches have different ways of knowledge representation methods.

3) Knowledge Base: The body of KBS mainly consists of the inference engine and KB, which exist as separate modules and work together closely [7]- [23]. They provide the main function of KBS.

The KB is a repository that stores all relevant knowledge acquired from domain experts through the knowledge acquisition process [7]- [23]. The acquired knowledge, which is structured and encoded in the knowledge representation process, is stored in the knowledge base. The structure of the knowledge base differs based on the selected KBS development approaches. Different KBS development approaches have different ways of structuring the knowledge in the knowledge base.

4) Inference Engine: The inference engine is an instruction on how the knowledge stored in knowledge-based is used for problem-solving [7]- [23]. It performs a search and deduces facts or draws conclusions from expert knowledge stored in the knowledge base. Inference engines accept end-user questions in a natural language form, formulate them to be understandable for a computer program, find the solution for the end-user questions and present the answer in the form of natural language. Inference engine uses reasoning mechanisms to manipulate, control, uses, and interpret expert knowledge and provides expert recommendations or advice [7]- [23].

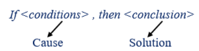
5) User Interface : Once the construction of the KBS is completed, it requires one critical module called the user interface to interact with system users simply by hiding KBS's complex internal structure [7]- [23]. The quality of the user interface(UI) determines the acceptance of KBS. To get user acceptance, the KBS UI must be user-friendly. KBS uses a graphical user interface(GUI) or natural language base question and answer, menu-driven user styles to implement the UI [7]- [23]. However, currently, most KBS use GUI. The user interface is where the KBS accepts end-user questions as input and displays experts' solutions for the end-user as an output.

IV. Knowledge Based System Development Approaches

Knowledge base system development approach, also known as knowledge base reasoning technique, is used to implement KBS or ES [7]- [23]. Knowledge base system development approaches differ on the way they are used to represent knowledge, on their knowledge base structure design, and their inference engine reasoning. The well-known Knowledge base system development approaches are rule-based, semantic networks or ontology-based, case-based, ANN Based, frame-based, object-oriented based and scripts based approaches [7]- [23]. For the purpose of this study, from the available reasoning mechanisms, the most commonly used three reasoning techniques; rule-based, case-based, and ANN-based techniques are selected and discussed as follows.

A. Rule Based Development Approach

Rule-based knowledge base system development approach, which is also known as rule based reasoning (RBR) technique is one of the most popular reasoning techniques that represents a piece of knowledge in the form of rules facts [24]. The rule is a conditional statement that derives a conclusion based on the fulfillment of a condition. The basic form of rule-based reasoning is the "IF-THEN " approach.



From the "IF-THEN" statement, the "IF" side expresses the premises or condition that needs to be true in order to execute the rule, and the "THEN" side of the statement states action needs to be taken or the conclusion needs to derive.

In the rule-based KBS development approach two inference methods are available. They are known as forward chaining (Data-Driven) and Backward chaining (Goal Driven) [25]. Both inference methods follow the same procedure and have the same function, they only differ in their inference technique and in the structure of the KB.

a) ***Forward chaining (Data-Driven):*** In the forward chaining inference method, the inference engine receives a problem as a condition and derives the conclusion that satisfies the condition. First, the system will crosscheck all the rules to find the one whose condition satisfies. So, with this scenario, a rule is executed at least once whether a fact that satisfied it is available in the database or not.

b) ***Backward chaining (Goal Driven) :*** This inference method only focuses on the rules that are applicable to a particular goal or conclusion. Unlike to forward chaining, backward chaining does not check rules that are not related to its conclusion or goal. Backward chaining receives a problem as a conclusion and tries to find premises or conditions that will cause the conclusion. In another expression, the system accepts a goal and then tries to find the fact that can match the goal.

B. Case Based Development Approach

Case-based knowledge base system development approach, also known as case based reasoning (CBR) technique, is an automated reasoning and decision process whereby it solves new problems through experts' experience accumulated in solving previous ones [26], [27]. CBR is built by storing the condition of the problem and the action taken at that time then the KBS focused on recognizing the similarity of new problems to existing previous ones. It is used for diagnosis, prediction, classification or recommendation of solution to the new case by remembering previous similar case [26], [27].

CBR represent a piece of knowledge in the form of cases. Cases, in CBR, can taken as a previous experienced situation collected and stored over time. CBR focuses on managing cases to solve a new case by finding previously recorded situation that are similar with the new one.

In order to build the KBS using the case-based development approach, knowledge must be acquired from previous situations and experts' solving experience first, then knowledge must be represented in the form of cases. The knowledge base and the inference engine should be designed based on the case-based cycle process. The user interface should also align with and support the case-based process.

C. ANN Based Development Approach

In the ANN based development approach, also known as ANN based reasoning, knowledge is represented as a structure of human brain systems [28], [29]. It is designed by making artificial neurons to be an approximately simulated model of human brain neurons. ANN has a large connected network, formed by neurons which takes an input and compute the desired output by processing the input values through multiple layers. ANN store the knowledge in the form of weight in which the value is decided based on the importance that particular input has to calculate the desired output. Some computation is taken place between input and output layers, in the hidden layers [28]- [30].

Layers are used to hold neurons and pass them to the subsequent layers. In the ANN based development approach neuron is an input point. Each neuron is a mathematical operation that takes its input to multiply it by the weight associated with it and passes the sum through the activation function to the other neurons.

As can be seen in fig 2 below ANN generally has three layers known as input layer, hidden layer, and output layer.

Input Layer: accepts all the inputs provided by the user

Hidden Layers: are set of layers that can be exist between input and output layers.
: perform some computations which results in the output.
: results are gained from this layer.

Output Layer: is the layer where the output is delivered.

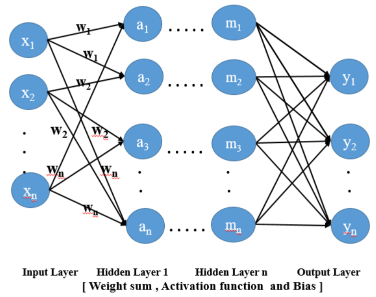


Fig. 2. Artificial Neural Network (ANN) model [30]

In each layer, there are a number of neurons that take the input and perform a mathematical operation by multiplying the input with weights associated with it and passes the sum through the activation function to the other neurons. The activation function computes the output from the weight of some of the inputs. It also normalizes the computed input to produce an output. Hence ANN deals with numeric data knowledge, knowledge needs to be converted from text to binary.

There are a number of ANN types. The most applicable ANN types are feedforward neural network and convolutional neural network.

*a) **Feedforward Neural Network (Artificial Neuron):*** In this ANN type, the input in the NN travels in one direction. It only used front propagation wave and backpropagation is not allowed in this ANN type. Feedforward NN is the simplest form of ANN, which makes it easier to be build and maintained. The input data is entered through the input layer nodes, the weighted sum of the nodes is calculated using classifying activation function and the output is exited on the output layer nodes. Such type of ANN may or may not have a hidden layer [28]- [30]. Feedforward neural network applications are found in speech recognition, computer vision, face recognition, and simple classification [28]- [30].

*b) **Convolutional Neural Network:*** A convolutional neural network encloses a three-dimensional array to arrange neurons, instead of having a two-dimensional array as standard. The first layer of a convolutional neural network is known as the convolutional layer. In the convolutional layer, neurons can only process from the small part of the input features by filtering them in batch-wise and computing this operation many times until the full visual image or the input is processed. The convolutional neural network is similar to a feedforward network, yet its neurons have learnable weight. Convolutional Neural Network applications are found in speech recognition, computer vision, video recognition, machine translation, semantic parsing, signal processing, image classification, paraphrase detection, and image processing [28]- [30].

V. Experimental Analysis

The overall experimental process of this research work is composed of four components such as data collection, data preprocessing, building knowledge base system model, and finally the model evaluation and comparison.

A. Data Collection

For this work, the 6-month incident log, extended from 01-09-2020 up to 28-02-2021 contains 4531 incidents extracted from IMARS. To complete the required knowledge, an interview is held with domain experts to collect the causes or indicators of the incident, which are used to identify the incident types and the possible expert solution and mainly required for this study. To simplify the work, the interview is done after data preprocessing with selected top 10 incident types. From the result of the interview and data analysis, 30 causes or indicators that are used to identify the selected top ten incident types are collected.

B. Data preprocessing And Attribute Selection

Incident types and features that are relevant for this study were selected using the data preprocessing process from the raw data collected from the IMARS incident log and through interview and data analysis. The data preprocessing process includes data reduction, data cleaning, data integration, and data transformation.

1) Data Reduction: Dimensionality reduction has two sections: feature extraction and feature selection.

a) Feature Extraction: The top 10 incident types, which are taken as a class label and dependent variable for the models, are selected by taking their frequency of occurrence as a major criterion. Applying this technique step enables to reduce the size of the dataset from 4531 incidents to a total of 1924 incidents.

b) Feature Selection: In this study, filter strategy is applied on the dataset, which contains the 30 causes or indicators of incident types as independent variables and the top 10 incident types as a dependent variable. Then the 30 incident causes or indicators of incident types were selected accordingly and preserved as features to model the KBS.

2) Data Cleaning: For this study, this process runs over the reduced dataset which contains 1924 total incidents. There were missed values that cannot be replaced by the probable values, accordingly, those values are removed from the dataset and a total of 1786 incidents are preserved as a complete and accurate dataset.

3) Data Integration: In this stage the data set, which is previously obtained from IMARS incident log, through interview and from data analysis preprocessed by passing through data reduction and data cleaning phases, integrated into one completed data set.

4) Data Transformation: In this process, the data set is transformed and coded into a format, which is suitable to model the rule-based, case-based, and ANN-based KBS development approach.

C. Building Knowledge base system model

The final data set used to build the KBS model contains 30 causes or indicators of incident types that are taken as independent variables and features, the top ten incident types are taken as dependent variables or class labels, and 1786 incidents taken as instances.

1) Rule-Based knowledge-based system model building:

a) Knowledge Acquisition: knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the rule-based KBS model.

b) knowledge Representation: For the purpose of this study, the preprocessed domain knowledge is changed into the form of "IF-THEN "rules. From the complete dataset, 32 rules are derived to predict or identify the class labels, which are the selected incident types. In addition, the rules are designed based on the class features, which are identified as causes or indicators of each incident type with the expert's solution.

c) Building Rule-Base knowledge-based system Model : The rule-based knowledge-based system model is built using the python programming language. The rule engine is designed as a KB based on the 32 rules derived in the knowledge representation step. The system is also designed and implemented to infer the value of the input features as causes and deduce the appropriate class labels, which are incident types, and recommended expert solutions by cross-checking the input regardless of the rules.

2) Case-Based knowledge-based system model building:

a) Knowledge Acquisition: knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the case-based KBS model.

b) knowledge Representation: The preprocessed domain knowledge is converted into cases or experiences. From the completed dataset, 32 cases are drawn based on the 30 features obtained from incident causes or indicators, to predict or identify the most similar class labels, which are the incident type with the appropriate expert solution.

c) Building Case-Base knowledge-based system Model: The case-based knowledge-based system model is built using the python programming language and based on the case-based cycle process. The case-based library is designed as a KB based on the 32 previous cases prepared in the knowledge representation step. The system is also designed to prepare the new case using the value of the input features and infer to deduce class labels, which are the incident types and recommended expert solutions, by crosschecking the new case with existing cases and suggesting the one with the highest similarity.

3) ANN-Based knowledge-based system model building:

a) Knowledge Acquisition: knowledge is acquired from domain experts through interviews and from data analysis. The collected knowledge is elicited and structured during the data preprocessing phase. The same data is used to build the ANN-based KBS model.

b) knowledge Representation: To represent the knowledge for the ANN-based KBS model, the preprocessed knowledge, which contains the 30 input features and the 10 class labels, is encoded from text variables into dummy variables. Dummy variables are variables that have zero and one values. The variable with the value of one indicates the presence or existence of that variable whereas the variable with the value of zero specifies the absence of the variable. Then the knowledge is coded accordingly to predict the ten class labels.

c) Building ANN-Based knowledge-based system Model: From the available ANN types, a feed-forward neural network that is suitable for classification is selected for this study. The model is built using the selected 30 features which are the causes or indicators of the incident types. So, the input layer has 30 neurons to handle the 30 features.

The number of the epoch, hidden layers, and neurons across each hidden layer affect the performance of the ANN-based KBS model. So ANN-based approach is built and trained using different numbers of epochs, hidden layers, and neurons. An epoch is a number of cycles used to train the neural network with all the training datasets. Then finally, the number of an epoch, hidden layers, and neurons that came up with higher performance is taken for this study. Therefore to build the model

- Two hidden layers are used to process the input features using relu as an activation function.
- 25 and 15 hidden neurons are used for the first and second hidden layers
- Out of the 1786 instances 70 percent of the data set is used to train the ANN-based KBS model
- Out of the 1786 instances 30 percent of the data set is used to test the ANN-based KBS model
- 200 epoch cycles are used to train the ANN-based KBS model

Hence the output of the ANN-based KBS model consists of 10 class labels, which are the top 10 incident types with their appropriate experts' solution, the output layer is designed to have 10 neurons and process the output using softmax as an activation function.

VI. RESULTS AND DISCUSSION

Knowledge base systems have different performance. The development approaches used in KBS inference engine for reasoning the conclusion and to construct the knowledge base engine affect the performance of KBS. This study shows how the performance of the KBS depends on the applied KBS development approach. The study compares the performance of KBS modeled build using rule, case, and ANN-based approaches. Performance evaluation metrics such as execution time, accuracy, F-measure, precision, recall, error rate, and MSE are used to compare the KBS development approaches.

The same dataset is used to test all KBS development approaches models. The data set contains 30 causes or indicators of incident types taken as independent variables and features, the top ten incident types taken as dependent variables or class labels, and a total of 1786 incidents taken as instances.

The result of the study is analyzed based on the outcome of the execution time taken by the system to respond to the input dataset, the accuracy, error in incident type identification using error rate, MSE, and F-measure including precision and recall.

Execution Time

As stated in table 5.1 and fig 5.1 unlike rule-based and case-based, ANN based KBS development approaches takes the lowest time, which is 0.95 seconds to build and execute the model to predict or identify the class labels or incident types with recommending expert solution. Rule-based and case-based KBS development approaches execute in 2.7 and 6.8 seconds respectively. Compared to case-based approaches, rule-based approaches execute with slightly minimum time.

KBS Development approaches	Execution Time (*in second*)
Rule Based Reasoning	2.7
Case Based Reasoning	6.8
ANN Based Reasoning	0.95

TABLE I
EXECUTION TIME TAKEN BY EACH KBS DEVELOPMENT APPROACHES

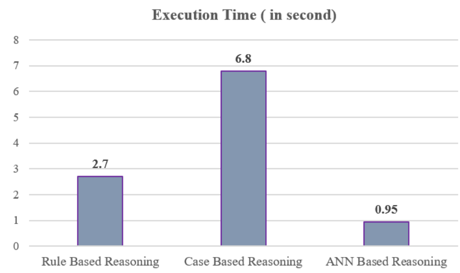


Fig. 3. Execution time metrics in second

Accuracy

An accuracy of 98.1 % achieved by the rule-based KBS development approach. Whereas the case-based and the ANN based development approach record accuracy of 96.8 % and 97.45 % respectively. In this result, the case-based development approach is the least performer. Compared to others, the rule-based KBS development approach has better accuracy in predicting or identifying the accurate class label or incident type and recommending the appropriate solutions. The result of the accuracy performance measurement is presented below in table 5.2 and fig 5.2.

KBS Development approaches	Accuracy (%)
Rule Based Reasoning	98.6
Case Based Reasoning	96.8
ANN Based Reasoning	97.45

TABLE II
ACCURACY PERFORMANCE OF EACH KBS DEVELOPMENT APPROACHES

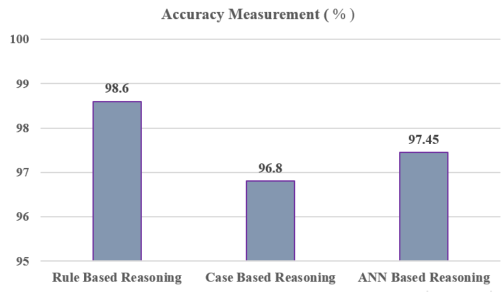


Fig. 4. Accuracy performance of each KBS development approaches

The F-measure, precision, and recall presented in table 5.3 and figure 5.3 below shows the same result as the overall accuracy can show in table 5.2 and figure 5.2. The result clearly shows that the rule-based development approach has the highest F-Measure, precision, recall. This indicates that the rule-based development approach is more efficient to correctly predict or identify the class labels or incident types and respective recommended expert solutions with less probability of miss predicting or identifying the class labels.

KBS Development approaches	F-Measure	Precision	Recall
Rule Based Reasoning	0.99	0.99	0.99
Case Based Reasoning	0.97	0.98	0.97
ANN Based Reasoning	0.97	0.98	0.97

TABLE III
F-MEASURE, PRECISION AND RECALL PERFORMANCES

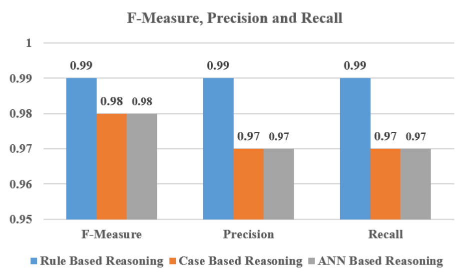


Fig. 5. F-measure, Precision and recall performances

Errors

The rule-based KBS development approach is the best approach in minimizing errors, hence it scores the minimum error rate and MSE which is 1.4 % and 0.187. Next to the rule-based approach, ANN based development approach has the minimum error rate and MSE which is 2.54 % and 0.35. Whereas the case-based development approach produces major errors in predicting or identifying the class labels or incident type and recommending an expert solution by recording a 3.19 % error rate and 0.49 MSE. The result of the error performance measurements is presented below in table 5.4 and fig 5.4.

KBS Development approaches	Error rate	Mean Square Error (MSE)
Rule Based Reasoning	1.4	0.187
Case Based Reasoning	3.19	0.49
ANN Based Reasoning	2.54	0.35

TABLE IV
ERRORS IN KBS DEVELOPMENT APPROACHES

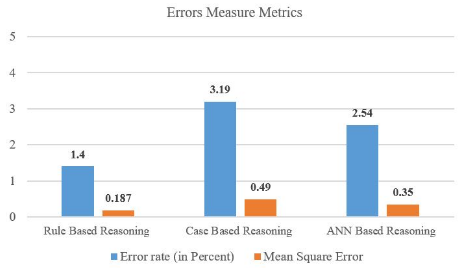


Fig. 6. Errors in KBS development approaches

VII. CONCLUSION

This study is conducted to identify the best KBS development approaches appropriate for the telecom operation and maintenance process, which has the best performance in accuracy, execution time, F-measure, precision, recall and produce less error in error rate and MSE. Rule-based, case-based, and ANN-based KBS development approaches are selected and compared on their ability in identifying class labels or incident types and propose an expert solution.

According to the result, the ANN-based KBS development approach achieves a better performance in execution time evaluation metrics by building and executing the model and presenting the result in 0.95 seconds. Rule-based and case-based KBS development approaches accomplish 2.7 and 6.8 seconds in the execution time respectively. In another way, the rule-based KBS development approach scores the best in accuracy measurement, which is 98.1 %. In these metrics the case-based and ANN-based development approach record 96.8 % and 97.45 %. In this result, the case-based development approach is the least performer. Compared to others, the rule-based KBS development approach has better accuracy in identifying the accurate class labels or incident types and recommending the appropriate solutions. As the same as the overall accuracy, the rule-based KBS development approach has the highest F-Measure, precision, recall performance. The rule-based KBS development approach also achieves the best performance in minimizing errors, hence it scores the minimum error rate and MSE which is 1.4 % and 0.187. ANN-based development approach produces 2.54 % error rate and 0.35 MSE. Whereas the case-based development approach produces a 3.19 % error rate and 0.49 MSE.

According to the result of the performance evaluation, the rule-based approach is the best KBS development approach in accuracy, error rate as well as in MSE evaluation metrics and ANN-based KBS development approach is the best KBS development approach to build a fast response time KBS.

In the ethio telecom context, the response time of the KBS is related to the handling time of incidents and gives a high priority. Accuracy of the KBS to suggest the exact class label or incident type and related expert solutions, also affect the handling time of an incident. If a high priority is given to response time, ANN-based development approach is a better solution to build a KBS for ethio telecom. While a high priority is given to accuracy and error metrics, the rule-based development approach is the better one. It is known that all evaluation metrics have an impact on the handling time of an incident and are taken as mandatory metrics. So, as a conclusion, this study proposes a hybrid KBS development approach by combining both rule-based and ANN-based KBS development approaches to be benefited from both approaches. The proposed approach enables ethio telecom to have a fast and accurate KBS with minor errors in identifying the occurred incident types and recommend expert solutions. Indirectly, the proposed approach allows ethio telecom to resolve its challenge occurred by the time its domain expert can't exist and enable to provide consistent operation support with available IT technicians by adopting expert solutions.

REFERENCES

[1] E. F. Vail, "Knowledge mapping: Getting started with knowledge management," Inf. Syst. Manag., vol. 16, no. 4, pp. 1–8, 1999.

[2] B. J. Bowman, "Building knowledge management systems," in IS Management Handbook, Auerbach Publications, 2013, pp. 877–892.

[3] B. Grasic and V. Podgorelec, "Developing knowledge management systems: Approaches, technologies and methods," in Proceedings of the 2011 American conference on applied mathematics and the 5th WSEAS international conference on Computer engineering and applications, 2011, pp. 207–212.

[4] E. M. Khaemba, "Factors affecting knowledge management implementation in telecommunication industry: A case study of safaricom limited," International Journal of Engineering Research and Technology, vol. 2, no. 6, pp. 1358–1366,2013.

[5] X. Huang and J. R. Jensen, "A machine-learning approach to automated knowledge-base building for remote sensing image analysis with gis data," Photogrammetric engineering and remote sensing, vol. 63, no. 10, pp. 1185–1193, 1997.

[6] L. Atymtayeva, K. Kozhakhmet, and G. Bortsova, "Building a knowledge base for expert system in information security," in Soft computing in artificial intelligence, Springer, 2014, pp. 57–76.

[7] E De Kock, "Chapter 6–expert systems and knowledge acquisition," University of Pretoria etd, pp. 107–125, 2003.

[8] M. D. C. Tongco, "Purposive sampling as a tool for informant selection," Ethnobotany Research and applications, vol. 5, pp. 147–158, 2007.

[9] H. Alshenqeeti, "Interviewing as a data collection method: A critical review," English linguistics research, vol. 3, no. 1, pp. 39–45, 2014.

[10] R. Burnashev, I. Amer, and A. Enikeev, "Expert system building tools based on dynamically updated knowledge," in Journal of Physics: Conference Series, IOP Publishing, vol. 1352, 2019, p. 012 008.

[11] A. I. Enikeev, R. A. Burnashev, and G. Z. Vakhitov, "Software tools and techniques for the expert systems building," in Fourth International Congress on Information and Communication Technology, Springer, 2020, pp. 191–199.

[12] E. Adetiba, T. John, A. Akinrinmade, F. Moninuola, O. Akintade, and J. Badejo, "Evolution of artificial intelligence languages, a systematic literature review," arXiv preprint arXiv:2101.11501, 2021.

[13] G. Neumann, "Programming languages in artificial intelligence," Encyclopedia of Information Systems, pp. 31–45, 2002.

[14] J. Sundnes, Introduction to Scientific Programming with Python. Springer Nature, 2020.

[15] B. Heinold, "A practical introduction to python programming," Creative Commons Attribution, 2012.

[16] L. B. Kaimal, A. R. Metkar, and G Rakesh, "Self learning real time expert system," International Journal on Soft Computing, Artificial Intelligence and Application (IJSCAI), vol. 3, no. 2, pp. 13–25, 2014.

[17] S. Mandal, S. Chatterjee, and B. Neogi, "Diagnosis and troubleshooting of computer faults based on expert system and artificial intelligence," International Journal of Pure and Applied Mathematics, vol. 83, no. 5, pp. 717–729, 2013.

[18] M. M. Mustaffa, A. Shibghatullah, A. Basari, and B Hussin, "Diagnosing computer hardware failures using expert system (rule-based technique)," in International Symposium on Research in Innovation and Sustainability, 2014.

[19] Y. Wang, Q. Li, M. Chang, H. Chen, and G. Zang, "Research on fault diagnosis expert system based on the neural network and the fault tree technology," Procedia Engineering, vol. 31, pp. 1206–1210, 2012.

[20] S. Bringsjord and B. Schimanski, "What is artificial intelligence? psychometric ai as an answer," in IJCAI, Citeseer, 2003, pp. 887–893.

[21] J. E. Aronson, T.-P. Liang, and R. V. MacCarthy, Decision support systems and intelligent systems. Pearson Prentice-Hall Upper Saddle River, NJ, USA: 2005, vol. 4.

[22] G. Yost and A. Newell, "A problem space approach to expert system specification.," in IJCAI, Citeseer, vol. 89, 1989, pp. 621–627.

[23] K. Tripathi, "A review on knowledge-based expert system: Concept and architecture," IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches and Practical Applications, vol. 4, pp. 19–23, 2011.

[24] J. Prentzas and I. Hatzilygeroudis, "Categorizing approaches combining rule-based and case-based reasoning," Expert Systems, vol. 24, no. 2, pp. 97– 122, 2007.

[25] B. Buchanan and R. Duda, Principles of rule-based expert systems. stanfrod, usa, 1982.

[26] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," AI communications, vol. 7, no. 1, pp. 39–59, 1994.

[27] A.-B. M. Salem, "Case based reasoning technology for medical diagnosis," World academy of science, engineering and technology, vol. 31, no. 2007, pp. 9–13, 2007.

[28] S. M. Saif, M. Sarikhani, and F. Ebrahimi, "An expert system with neural network and decision tree for predicting audit opinions," IAES International Journal of Artificial Intelligence, vol. 2, no. 4, p. 151, 2013.

[29] D. Arditi and O. B. Tokdemir, "Comparison of case-based reasoning and artificial neural networks," Journal of computing in civil engineering, vol. 13, no. 3, pp. 162–169, 1999.

[30] G. Liu, X. Yang, Y. Ge, and Y. Miao, "An artificial neural network-based expert system for fruit tree disease and insect pest diagnosis," in 2006 IEEE International Conference on Networking, Sensing and Control, IEEE, 2006, pp. 1076– 1079.