Supervised learning algorithm's role in flagging programming concepts that call for Information Technology teachers' attention

By

1672247

Mashite Tshidi

Research Report submitted in fulfilment of the requirement for the **MASTER OF EDUCATION** degree

In the

Faculty of Humanities

EDUCATIONAL INFORMATION AND ENGINEERING TECHNOLOGY (EDIET) DIVISION

UNIVERSITY OF THE WITWATERSRAND

Supervisor: Dr Alton Dewa

Date: 30 November 2022

Declaration

I, Mashite Tshidi, certify that the work titled "**Supervised learning algorithm's role in flagging programming concepts that call for Information Technology teachers' attention**." is unique to me and has not previously been published or offered for publication anywhere. Since other people's work has been used, it has been appropriately acknowledged in the document, and a detailed, organised reference list has been supplied. I am aware that if it is discovered that this was not my original unassisted work or that I had not adequately cited any of the ideas or phrases I used, the University of the Witwatersrand will take disciplinary action against me.

Name: Mashite Tshidi

Signature: 

Date: 30 November 2022

Dedication

This research report is dedicated to the woman who raised me, my grandmother, Nkele Josephine Tshidi. Thanks for the valuable life teachings and for introducing me to God; I am who I am today because of you.

Acknowledgements

First and foremost, I want to thank God for wisdom, fortitude, and the opportunities He gave me to complete this paper. I will always be grateful. To those who came before me, thank you for supporting me along my path.

My sincere appreciation is extended to my supervisor, Dr Alton Dewa, for his ongoing assistance with my education and patience during my study. The possibility of this research would not have existed if it were not for your commitment and extra work in monitoring me. This study report was engaging and amusing, thanks to you.

My parents, Josephine Tshidi and Isaac Mankge deserve more gratitude than I can ever say. I appreciate your unwavering support and love.

I extend gratitude to my academic tag team partners, Ms Ndileka Mnqonywa and Dr Hlamulo Mbhiza. To Dr Mbhiza, your mentorship has impacted me beyond academics, and I am internally grateful for you. To my infield study buddy and friend, Ndileka, we share our wins and losses. Thank you very much for your assistance and insightful feedback during this research.

To my therapist at Ollie Health and one of my closest friends, Ms Gugu Mthimkhulu, thanks for lending me an ear. Your assistance in stabilising my mental health and finding healing during the losses and heartbreaks I faced while completing this report was instrumental.

I am so thankful to my friends and work colleagues. A sincere appreciation to Ms Emily Ngwenya who played a crucial role in selecting the topic of study for this research report. To my colleagues at Valenture Institute and Wits School of Education thank you for making work feasible whilst providing comprehensive support in my academic journey.

Thank you to the National Research Foundation (NRF) for financially supporting my Master of Education studies. Without your scholarship, it would not have been viable for me to finish this study.

The Information Technology teachers who volunteered their valuable time to share their views are much acknowledged for rendering the study possible; without you, this study would not have been achieved.

Lastly, to me, thanks for showing up.

Abstract

Programming syntax and concepts' complexity demands learners possess logical thinking and problem-solving skills to effectively write and complete code. In the context of Information Technology education in South Africa, teachers require a tool that can identify learners' performance in programming concepts and help them prevent misunderstandings. This research proposes a transformational approach that uses a machine learning algorithm to alert teachers of programming concepts that learners may struggle with. The study also investigates how Information Technology teachers shape learning experiences when teaching programming concepts, using a qualitative methodology involving semi-structured interviews with Information Technology teachers. The study employs Educational Data Mining and Learning Analytics as theoretical and conceptual frameworks to showcase the potential of supervised learning algorithms in using prior Information Technology results for significant improvements in learning and performance. The findings indicate that problem-based learning is a commonly used methodology among Information Technology teachers. The algorithm results reveal a high-performance forecasting model based on acceptable accuracy, actual positive rate, and false positive rate. The identified programming concepts that require focus include conditional statements, conceptualizing problems and designing solutions, debugging and exception handling, abstraction/pattern recognition, and differentiating between classes and objects. Overall, this research presents a valuable approach for leveraging a supervised learning algorithm to enhance Information Technology education by identifying and addressing programming concepts that learners struggle with.

Keywords: Programming, Information Technology, Machine learning, Educational Data Mining, Learning Analytics, Problem-based learning, Algorithms, Forecasting model.

Table of contents

List of Abbreviations

4IR – Fourth Industrial Revolution

AI – Artificial Intelligence

ATP - Annual Teaching Plan

CAPS - Curriculum and Assessment Policy Statement

DoE – Department of Basic Education (Republic of South Africa)

EDM – Educational Data Mining

FET – Further Education and Training

FP – False Positive Rate

GUI – Graphical User Interface

ICT – Information Communication Technology

IDE – Integrated Development Environment

IT – Information Technology

LA – Learning Analytics

NCS - National Curriculum Statement (Republic of South Africa)

OOP - Object-Oriented Programming

PAT - Programming Assistance Tools

PAT – Practical Assessment Task

PBL – Problem-based Learning

POPIA - Protection of Personal Information Act

SAGG – South African Government Gazette

STEM – Science, Technology, Engineering and Mathematics

TP – True Positive Rate

List of Figures

List of Tables

Chapter 1: Introduction

"If you want to understand today, you had to search yesterday." – Pearl Buck.

1.1 Background of the study

Information Technology (IT) is one of the 29 subjects taught in secondary schools from grades 10 to 12, as stated in the Curriculum and Assessment Policy Statement (CAPS) document. The subject was established in South Africa as a school subject that concentrates on problem-solving and incorporates Information and Communication Technology (ICT) with socioeconomic applications to participate in the global marketplace (DoE, 2008). The justification for this choice was to train a workforce with the abilities wanted to optimise software development and technical assistance (Havenga and Mentz, 2009). Considering the rise in businesses creating digital, mobile, and cloud-based solutions, ICT-related vocations have been globally recognised as leading vocations (Koorsse *et al.*, 2015). In addition, the authors point out that considering the high demand for ICT skills and experience, countries worldwide, including the United States, Europe, and South Africa, have seen a drop in computer science programmes at tertiary-level education institutions. This decline in enrolment has reduced the number of software developers available in the industry. Globally, countries have recognised the importance of ICT professions for the country's technological and economic growth (Wilson *et al.*, 2010). The South African curriculum has also acknowledged the significance of promoting ICT-related skills and knowledge in schools.

South Africa is experiencing an increased demand for IT graduates (Jacobs and Sewry, 2010). Schools have limited access to computer facilities, learners are less inclined to take the IT subject. Such declines and lack of interest raise concerns about whether the tendency will continue and cause problems in the employment and educational sectors (Seymour *et al.*, 2005). To circumvent these challenges, the South African education system outlines the learning objectives, outcomes, and essential tenets of teaching and learning in all subjects, including IT. The CAPS endorses the National Curriculum Statement (NCS) by establishing the groundwork for attaining a range of objectives. The attainment of these objectives aims at producing learners who can (DoE, 2011):

1. Use creative and critical thinking to recognise issues, find solutions, and make judgments.
2. Work well both alone and with others as team members.
3. Plan and manage their actions responsibly and successfully.

4. Gather, organise, analyse, and critically evaluate information.
5. Efficiently communicate in numerous ways, employing words, symbols, and visuals.
6. Utilize science and technology critically and effectively.
7. Furthermore, acknowledge that various contexts for problem-solving do not exist independently to demonstrate an understanding of the world as a system of related processes.

In the context of IT, the latter objectives are mirrored specific aims on the type of learners the curriculum envisages in alignment with the topics covered in the subject. The critical components of the goals can be depicted within the skills of IT encompasses activities for solving problems that require communication, information management, and logical reasoning (Koorsse *et al.*, 2010). Computer application development using existing object-oriented, net-centric software packages using programming components is another crucial aspect of IT that requires attention. (DoE, 2008). According to the DoE (2011), under the CAPS curriculum, an IT learner will:

- Plan solutions through applicable techniques and procedures.
- Comprehend and employ relevant communication tools for information distribution.
- Recognise the various system technologies used to build computer-based platforms.
- Acknowledge that software engineering fundamentals build the cornerstone of all ICT components.
- Understand using internet technology for a variety of purposes.
- Learn how a knowledge-driven society works by understanding while using data and information management techniques.
- ICT usage should be acceptable, and learners should know the societal ramifications of these tools.

The aims of the subject span six topics in the curriculum of IT subject. The topic areas covered in the issue include Solution Development, Communication Technologies, Systems Technologies, Internet Technologies, Data and Information Management, and Social Implications (See Table 1).

Table 1: Information Technology topic areas in Grades 10 – 12 (DoE, 2011)

Topic Area	Sub-Topics	Weighting (Content)	Resources
Solution Development	Algorithms and Problem Solving Introduction to Solution Development Application Development Software Engineering Principles	±60%	Computers Textbook Software • Database Management Software • High-level programming language within a visual development environment using an IDE with a GUI builder • Internet • Browser
Communication Technologies	Networks E-communication	±7%	
Systems Technologies	Introduction to Computers Hardware Software Computer Management	±10%	
Internet Technologies	Internet World Wide Web Internet Services	±8%	
Data and Information Management	Data Representation Database Management Database Design	±10%	
Social Implications	Legal Issues Ethical Issues Social Issues Environmental Issues Health Issues Computers and Society	±5%	

It is essential to note from Table 1 that there is topic overlinks and overlaps with the subject, which gradually leads to Solution Development. According to DoE (2011), solution development is enabled by system technologies in software applications. Electronic communication is possible thanks to system technologies. Electronic communication systems, used for various tasks, including exchanging information and data delivery, enable the Internet. The concepts of solution development and Internet technology share similarities with those of data and information management, both a fundamental premise and a secondary activity. System technologies offer data and information management. Many ICT operations are motivated by human engagement, need, and action, which raises social and ethical implications. This hierarchical representation informs the context in which this study, which is an endeavour to IT teachers about programming concepts that they need to pay attention to, which fall under Solution Development and carry the most weighting in terms of content and the weighting (See Table 1 above for the weighting of content and Figure 1 as the topic areas hierarchy in the subject).

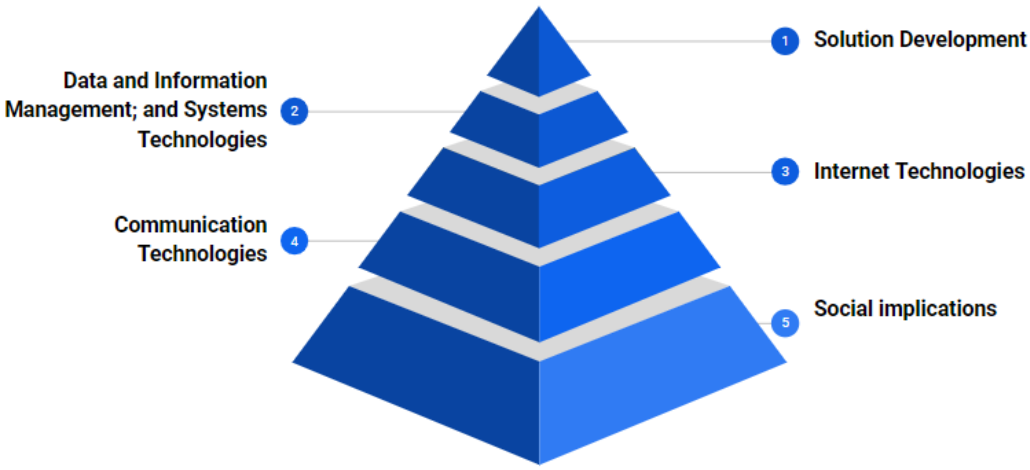


Figure 1: Hierarchical representation of the topic areas in IT - Grade 10-12

The method of producing programmes in a planned and ordered manner, known as Solution Development," primarily focuses on the logical problem-solving of computational issues. It entails developing methods and programmes using guidelines and specifications provided by a customer, business, or individual or by the problem description itself (DoE, 2011). A programming language is a language that sends programme code—or instructions—to a computer to conduct specific tasks. A user-defined or built-in object method is used while writing computer code using an object-oriented programming language (Weisfeld, 2009). Programme code is typed following the language's syntax and semantic rules in an object-oriented programming language (Kelleher and Pausch, 2005). This means appropriate processes, tools, and procedures must be used to construct the programme. Therefore, computer programming is used to create programmes. According to DoE (2011), it can be based on a single or a combination of development paradigms such as event-driven programming, object-oriented programming, and sequential programming. Figure 2 provides a layout of which target areas of these paradigms are for learners to be taught the fundamental programming principles and constructions using visual programming tools. Implementing a Graphical User Interface (GUI) in the enrolment of the development paradigms is supported by visual programming languages, with each programming example shown using graphical objects. It is stated that because of its capacity to facilitate forward and backwards thinking, stimulate memory, and visually represent a programme's control and data flow. Visual programming was easier to use for novices (Tijani *et al.*, 2020).

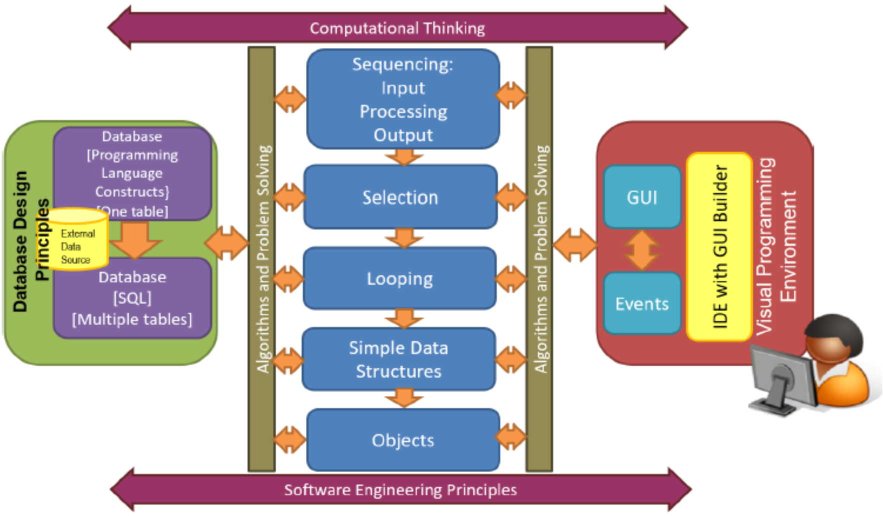


Figure 2: A comprehensive presentation of programming concepts in CAPS (DoE, 2011)

In 2012, under the curriculum transition between the NCS to CAPS, South African IT curriculum developers shifted in terms of the use of visual programming tools that were used in the subject from Scratch to Delphi. Scratch, an introductory programming language, was initially taught in Grade 10 before learners could progress to Delphi in Grade 11. According to van Zyl *et al.* (2016), this choice increased accessibility for learners and encouraged them to study IT in school. Subsequently, learners had to expand their Scratch programming knowledge to complete object-oriented programming in Delphi in grades 11 and 12. The motive of Scratch as an introductory programming language could improve learners' understanding of Delphi behind the curriculum developers for using dual programming languages. Scratch was created to inspire learners to let programming, foster organised and imaginative thought, and foster teamwork (van Zyl *et al.*, 2016). The Scratch environment is aesthetically appealing, and programming is accomplished by creating vibrant code blocks. Scratch is an object-based language rather than an object-oriented language. Due to curriculum changes, Delphi's visual programming tool is commonly adopted in South African schools.

Before programming can be done effectively, it is necessary to comprehend the language's logic, syntax, and general programming ideas (van Zyl *et al.*, 2016). Delphi, compared to Scratch, is a hybrid language for programming that allows for both conventional and object-oriented programming (Weisfield, 2009). Errors are frequently committed because strict syntax rules must be followed while generating the code (see Table 2).

Table 2: Programming concepts that need to be taught in Delphi

Programming concepts	Explanation of the content that needs to be covered based on programming concepts in Delphi between Grades 10-12 that learners need to be taught.
1. Problem-solving	● They are designing algorithms. ● Know how to build the interface and write the code. ● Correct syntax errors and run the application.
2. Integrated Development Environment (IDE)	● A GUI contains components and properties comparable to those in a Windows environment.
3. The programme's implementation	● By using a menu, programmers should place items for input, output, and processing on a form. In addition, learners must write code to link events to components.
4. Syntax	● A proper understanding of the Delphi syntax is required.
5. Variables and data types	● Programmers need to adhere to the basic rules of naming variables. ● Learners need to know that when programmes are running, variables are not displayed. ● Programmers need to know that there are several data types, and each must be treated differently. Each data type has a unique set of values. ● During the programme-built, users need to be familiar with the conversions of the data types often.
6. Programmes execution	● In Delphi, a compiler converts a programme into machine code. Learners need to be able to use the compiler to rectify syntax errors in a programme to generate an executable code.
7. Decision and looping structures	● An in-depth understanding of choices and looping structures is essential for writing programmes.
8. Object-oriented programming	● Delphi is an object-oriented programme. Therefore, all characteristics of an object-oriented language are available, including abstraction, polymorphism, inheritance, and encapsulation. ● Learners must be taught how to employ a procedural or object-oriented approach in executing their programmes.
9. Error handling and testing	● While implementing programmes, syntax, logic, and execution errors can occur. ● Learners need to be familiarised with the constructed debugging function that allows the programmer to go through a programme and inspect the contents of variables.

As a result of Table 2, the subject fosters computational thinking practices within programming, such as algorithm development and problem-solving, through grades 10 to 12. According to DoE (2011), learners are exposed to basic computational concepts and capabilities in Grade 10 and the ability to develop algorithms and solve problems. They use an IDE and a GUI builder to write code in a moderate programming language. The fundamentals of interfaces, coding, and event handling are presented to the learners. The concepts and ideas of programming learnt in Grade 10 are expanded upon in Grade 11. With the help of the form class, attributes, methods, and controls, event handling principles are highlighted. In Grade 11, learners also learn how to handle databases using coding techniques. Through more complex concepts and problems, the principles and constructions are further emphasised in grade 12 through the fundamentals of object-oriented programming that are reiterated to the learners (DoE, 2011).

Programming is a process that could be regarded as complex, particularly for a novice programmer employing Object-Oriented Programming (OOP) as Delphi. Before beginning to code, novice programmers must become familiar with the IDE (Koorsse *et al.*, 2015). A further step in programming entails selecting an object from a list of options before typing programme code for specific events connected to that object. Error warnings appear if code is not entered precisely by the programming language's syntax requirements (van Zyl *et al.*, 2016). Programme execution is prevented. As a result, addressing errors keeps programmers always working (Koorsse *et al.*, 2015). It can be an arduous process, and it takes some time for inexperienced programmers to comprehend these warnings and be able to fix the errors.

Programming difficulties are solved using a variety of strategies and techniques. A strategy describes the anticipated course of action for handling the programming issue and outlines how to move on with the problem-solving process. In this context, the ability to apply problem-solving thought processes to resolve a programming problem is referred to as an activity. When teaching programming-related material, it is frequently believed that teachers automatically use the necessary knowledge, skills, and methodologies to teach complex problem-solving tasks. Giving people adequate exercise will only assist if they are taught how to solve problems correctly and methodically (Hasni and Lodhi, 2011). Furthermore, Rist (1996) asserts that a lack of planning is the primary contributing factor to learners' difficulty with programming. To address this factor, the following issues need to be addressed: learners' incapacity to apply program components, lack of analytical abilities, inefficient use of problem-representation approaches, and inefficient use of problem-solving and coding instructional tactics (Ismail *et al.*, 2010).

Computer programming requires science (reasoning, problem-solving, and critical thinking) and art (creative thinking, program design, and development). Therefore, teachers must teach a variety of thought processes to support learners in their work. It is necessary to impart knowledge of interpretation, problem-solving, logic, programme design, synthesis, assessment, and reflection (Rist, 1996; Ismail *et al.*, 2010; Hasni and Lodhi, 2011). These actions are a component of procedural knowledge that explains "how" to move forward with the solution. Ismail *et al*. (2010) stress that learners must develop critical thinking and problem-solving skills before applying and using different tools and programming languages. Therefore, explicit problem-solving and programming instruction should focus on approaches to organise, represent, design, and resolve the current problem. This gives the premise of the topic area of Solution Development in IT. Detailed thought processes, problem analysis, and explicit instruction of problem-solving activities are expected to enable effective programming under the topic area and sub-topics.

Subsequently, with the advent of computer programming, IT is perceived as a thorny subject in secondary schools in South Africa. The subject is usually tricky because it requires sophisticated problem-solving abilities to effectively create computer programmes and specialised resources to teach it (Mentz *et al*., 2012). This reputation influences the decision to enrol in IT classes at school. According to Koorsse *et al*. (2015), before Grade 12, many learners who attempt the subject move to another, more basic subject. However, those who remain in the subject still require additional motivation and passion for the programming scope to succeed. Lack of enthusiasm for the subject and proficiency in programming adversely influence the rate of learners who decide to pursue higher education in a computing area (Koorsse *et al*., 2015).

The learning environment in secondary schools worsens these issues (Koorsse *et al*., 2010). As a provision by the DoE (2011), the school needs to supply the infrastructure and equipment for the subject. In addition, learners in IT must work independently on a computer throughout class time and require internet access. Nevertheless, most public South African schools need to be digitally savvy. In hindsight, the latter claim signifies that the influence of the apartheid system persists and has an impact on the South African educational system. Since public schools cannot levy fees to significantly boost resources, especially computer labs, learners in impoverished communities, such as rural areas and formerly Black urban townships, proceed to be academically disadvantaged (Galpin and Sanders, 2007). As a result, many learners still need to be exposed to computers in the classroom and are not allowed to choose IT as one of their FET subjects. In a technologically rich culture, learners can integrate new knowledge by expanding on the knowledge imparted by their teacher, resulting in more significant learning (Mukuna and Aloka, 2020). However, constraining access to educational resources lowers learner performance as they cannot create understanding.

Pears *et al*. (2007) claim that a dearth of instructional materials contributes to learners' worse performance and reduced willingness to engage in IT classes. IT learners also require extra resources to supplement the typically required IT subject texts. One kind of value addition is programming assistance tools (PATs). PATs are designed for novice programmers (learners) to use to assist them in comprehending programming concepts and developing programming skills (Koorsse *et al*., 2015). A PAT's programming is easier to use than professional programming environments and supports programming learning through visualisation techniques. Visualising programming concepts assists novice programmers in comprehending abstract concepts by leveraging engaging microworlds to make programming more exciting and applicable (Pears *et al*., 2007). A PAT has the advantage of supporting a novice programmer in gaining knowledge of programming concepts, providing an automated assessment of programming tasks, and correcting simple errors (Koorsse *et al*., 2015).

IT learners experience the same difficulties with programming as novice coders (Havenga and Mentz, 2009). Given that daily classes are typically 45 minutes, it is hard to master and practice sophisticated programming topics within the restricted school period (See Table 3).

Table 3: Approximate teaching time per topic (DoE, 2011)

	Grade 10		Grade 11		Grade 12	
Topic	**Hours**	**Weeks**	**Hours**	**Weeks**	**Hours**	**Weeks**
Solution Development	92	23	90	22.5	68	17
Communication Technologies	4	1	8	2	4	1
Systems Technologies	16	4	10	2.5	10	2.5
Internet Technologies	14	3.5	6	1.5	4	1
Data and Information Management	8	2	18	4.5	8	2
Social Implications	6	1.5	8	2	6	1.5
Teaching Time: Total	**140**	**35**	**140**	**35**	**100**	**25**
Examinations	**20**	**5**	**20**	**5**	**48**	**12**
TOTAL:	**160**	**40**	**160**	**40**	**148**	**37**

As a result, IT teachers need more time to address the unique programming issues each IT learner is experiencing or gauge each learner's conceptual comprehension during class (Koorsse *et al.*, 2015). This leads to the restructuring of the teaching strategies for the teacher to meet the individual learning needs of their learners based on their performance on assessments. The critical question is, "How accurately can a teacher assess how much of the intended material the learners have learned? Can IT teachers, for instance, reliably estimate how each learner would do on a forthcoming programming exam?" Of course, such determinations would constitute only a tiny portion of all the actions the teacher could engage in to assess learner learning. However, if teachers can make these conjectures correctly, it would be a crucial skill to help them better start regulating their learners' knowledge acquisition (Thiede *et al.*, 2015). Discovering ways to increase the accuracy of teachers' assessments of learner learning is particularly crucial in this era of increased responsibility, which is seen in several countries (Alenezi and Faisal, 2020), including South Africa. Subsequently, teachers are expected to make more accurate assessments of their learners' understanding and to use their diagnostic tests to continue supporting them during instruction. In alliance with machine learning and data mining techniques, these proactive interventions can aid teachers in identifying topics that learners are struggling with and help them get back on track (Coleman *et al.*, 2019).

Schools attempt to determine learners' achievement. Determining learner performance can aid educators in preventing learners from dropping out of a subject and recognising those who need proper assistance (Berland *et al.*, 2014). However, assessment is usually done at the end of the term, leaving the teacher with inadequate time to recognise learners who need assistance. Learner assessment is a necessary process in the educational sector. Educational assessment assesses a learner's degree of knowledge and enhances their learning throughout the course (Alenezi and Faisal, 2020). Educators are responsible for maintaining a record of their learner performance and the issues they experience. Using computer programmes, educators can readily forecast the obstacles learners confront in their academic procedures as technology improves in the twenty-first century (Alenezi and Faisal, 2020). Monitoring and directing learners through the instructional process and assessment are all part of gauging learners' performances. Assessments, the primary method of determining learning outcomes, reveal the degree of learners' performance, both subjectively and numerically. Therefore, aiding teachers during this procedure is crucial as it impacts learners' motivation, growth, and learning strategies and is regarded among the major entities in education (Bhutto *et al*., 2020).

Assessment is done at the end of each term in the learning process (DoE, 2011). This puts much strain on the IT educator to determine what obstacles have accumulated over time and what must be taught again (Bhutto *et al.*, 2020). This prompts the idea that feedback needs to be more adequate. Assessment feedback is essential in determining future learning because it directly impacts learners' performance and constant effort in upcoming tasks (Alenezi and Faisal, 2020). This study aims to demonstrate the importance of using a supervised learning algorithm to recognise learners' subjects or challenges, which teachers can confront early in teaching and learning. Human-understandable rules are represented by supervised learning algorithms (Sarker, 2021) employed in knowledge systems such as databases. The supervised learning algorithm is used to classify problems (Sarker, 2021) and present some solution indicators to the IT teacher.

1.2 Problem statement

The swift and parabolic proliferation of educational data and the issue of deploying that data to improve the quality of the education system are two major concerns that education institutions are presently facing (Berland, 2013). Consequently, these concerns encountered by learners can be addressed by proper educational data analysis, which includes having findings or testing hypotheses or algorithms on a dataset (Berland *et al.*, 2014). A study by Berland (2013) notes that teachers need a tool that indicates or points to the learners' need to understand the concept. Subsequently, a tool that determines learner performance in programming concepts is required to aid IT educators in preventing learners from misunderstanding the concept.

The tool should recognise learners who need immediate assistance and inform teachers to help them improve the misunderstood concepts. Islam, Mouratidis and Mahmud (2021) noted that a system can fill the gaps by combining current learner data to produce functional predictive and prescriptive analysis. This research attempts to present a novel supervised learning that integrates learner performance in programming and uses Learning Analytics (LA) to determine areas of programming that learners are struggling with. These areas in which learners struggle are sent to the teacher to help the learner. Machine learning algorithms can be used to support LA by creating models that can predict learners' outcomes based on their performance datasets (Sarker, 2021). The performance of a machine learning algorithm degrades when all attributes are used. Therefore, carefully selecting predictive features improves the model's performance (Khan *et al.*, 2021). The study proposes developing an algorithm that flags common misconceptions that learners face in summative assessments that teachers should be aware of in IT.

1.3 Rationale of the study

The rationale for conducting this study is to investigate the role of supervised learning algorithms in alerting IT teachers to programming concepts that their learners are struggling with. Webb *et al.* (2020) note that there is a need to accurately gauge learners' performance at different points throughout the school term, and knowledge discovery through machine learning algorithms can help teachers manage their classes more efficiently, pinpoint learning challenges, and strengthen their teaching approaches. This study aims to introduce teachers to data interpretation and statistical analysis from an algorithm, thus requiring them to make inferences on programming topics that learners are struggling with.

The research interest in supervised learning algorithms stems from the need for Education 4.0 tools within the 4th Industrial Revolution (4IR). Education 4.0 is an educational approach associated with the 4IR and intends to transform the future of education by using innovative technology and automated processes (Anelka, 2018). This is exacerbated by Chen *et al.* (2020), who note that in Education 4.0, the ability to measure the learners' performance lies at the heart of educational institutions. Subsequently, in Education 4.0, there is a need for tools that alert teachers immediately on concepts their learners are facing challenges with, to optimise learning experiences of those concepts. Considering South African schools are still traditionally based (van Zyl, 2016), Education 4.0 tools like machine learning algorithms can still be applicable to maximise the learning experiences (Anelka, 2018). Hence, this study aims to bridge the gap in research by investigating the potential benefits of supervised learning algorithms in improving programming education in the IT subject.

1.4 Purpose of the study

This study has three distinct purposes. The first one seeks to objectively analyse and examine IT teachers for Grade 10 on how they mould learning experiences when they instruct learners on programming concepts. The second one is to provide literature on established tools based on machine learning that can be used to let teachers understand what topics or concepts their learners are having difficulty with so they may step in and help. The basis of the preceding purposes informs that the third purpose of the reach is to construct a model and implement a supervised learning algorithm that identifies learners' programming difficulties. Due to its integration of Data Science and Education, this research is a component of a cross-disciplinary research endeavour. By integrating these two academic disciplines, it is possible to advocate a transformative strategy (Marshall and Geier, 2020) in which machine learning algorithms may be used to notify teachers of programming topics that learners are having trouble understanding and provide them with the chance to help.

1.5 Aims of the study

This study aims to investigate the potential of a supervised learning algorithm as a tool for identifying programming difficulties among learners in Grade 10 IT classes. To achieve the aim of the study, the following objectives are pursued:

- Review existing literature on machine learning algorithms and their applications in educational settings, particularly in identifying learner difficulties with programming concepts.
- Conduct semi-structured interviews with Grade 10 IT teachers to explore how they facilitate learning experiences for programming concepts.
- Collect and analyse data on learners' recent exam results to determine their performance in programming concepts.
- Develop and implement a supervised learning algorithm that identifies programming concepts that learners are struggling with.

1.6 Research questions

The primary research study question is:

What tool do IT teachers require to be notified of programming concepts in which learners are having difficulty and to intervene?

The following supporting research questions were identified to further the primary research study's question:

1. How and when do teachers intervene with learners facing programming difficulties?
2. How effective is the method in preparing learners' understanding of the concept?
3. What effective tool informs IT teachers about issues in which learners are having difficulty?

1.7 Structure of the research report

There are six chapters in this research report. The primary objective of **Chapter 1** is to establish the study's background by emphasising the characteristics and requirements of the IT subject. This leads to an account that discusses the issue of the notion that programming is a challenging endeavour. This research suggests a transformative strategy in which machine learning algorithms could be used to notify IT teachers of programming concepts learners are having trouble understanding and give them a chance to act at an early convenience. This chapter also contains the justification for the study's conduct, purpose, and research objectives and questions.

Chapter 2 presents literature that focuses on machine learning algorithms. This chapter examines the application of machine learning as a concept in educational activities. This is done by looking at several methods in which teachers apply machine learning technology in their classrooms. This study delineates the range of learning alternatives supplied by machine learning for an educator to choose from various pedagogical strategies and technological tools that can amplify proficiency in programming techniques in consideration of the individual differences between learners through the advent of virtual assistance. This chapter also looks at how teachers can use the alliance between machine learning and virtual assistance to track their learner's understanding of a concept through their learning activities at any given time. With a focus on supervised learning algorithms, the chapter discloses how this automated tool can be used to improve learning experiences. Subsequently making it easier for teachers to raise their standard of instruction.

Chapter 3 discusses a presentation on the theoretical and conceptual frameworks espoused to explore programming concepts teachers need to pay attention to. To achieve this in the light of this study, Educational Data Mining is sought as a theoretical framework, and Learning Analytics aids it as a conceptual framework to enhance the prominence of the study findings.

The scope of this investigation is described in **Chapter 4**. Detailed information about the study's chosen research methodology is given. The reader will learn how this study uses a qualitative research methodology and techniques in this chapter. Moreover, this section outlines the sampling methodology of the study by glancing at the sample size's justification. From there onwards, the issues of credibility, transferability, reliability, and confirmability are foregrounded.

In **Chapter 5**, the results and outputs from the supervised learning algorithm are presented in conjunction with the discussions from the interviews conducted with IT teachers. These results, outcomes and discussions are then examined considering the literature review and the established theoretical and conceptual framework for the investigation.

In **Chapter 6**, the conclusions are discussed concerning the theoretical and conceptual frameworks, research objectives and questions, and reviewed literature. The importance of the study, considering the gaps found in the current literature, is also highlighted. The study's limitations are also discussed. The chapter concludes by outlining prospective opportunities for future research.

Chapter 2: Literature Review

"*We are entering a new world. The technologies of machine learning, speech recognition, and natural language understanding are reaching a nexus of capability. The result is that we will soon have artificially intelligent assistants to help us in every aspect of our lives.*" – Amy Stapleton.

2.1 Chapter overview

The ability to collect enormous amounts of data has been facilitated by the quick advancement of the Internet and related technologies (Aalst, 2016). However, these vast amounts of data frequently result in information overload. Information overload occurs when a person's cognitive abilities cannot handle the quantity of input (for example, data) they are trying to comprehend (Injadat *et al*., 2021). Subsequently, people who are overloaded with information may ignore, neglect, or misconstrue essential details. The cognitive ability of humans to process massive volumes of data is not present. Data science has therefore become a field of study (Injadat *et al.*, 2021). To extract information from massive datasets, data scientists integrate the traditional areas of statistics, data mining, databases, and distributed systems (Aalst, 2016). Machine learning is one kind of data analysis that data scientists can employ. Machine learning enables computers to learn without explicit instructions. Once the computer has identified patterns in a learning dataset, it may use what it has discovered to place them in different datasets (Sarker, 2021). Therefore, machine learning enables computer systems to change and improve over time. However, the applicability of the evolution of these computer systems can result in the transformation of education needs to be studied. This chapter starts by outlining the most recent advances in machine learning in the 4IR. The relevance of machine learning to education is inferred from that description. As a result, the debate about the advantages machine learning may bring for Education 4.0 is framed around the models it offers. A thorough explanation of classification and decision-tree supervised learning techniques illustrates how programming concepts are identified by applying these algorithms. Finally, a review of research highlighting the study gaps is done.

2.2 Understanding machine learning

Machine learning is a broad field with implications in IT and many other areas. In AI, machine learning combines statistics and computer science to create more effective algorithms when presented with relevant data instead of providing explicit instructions. In addition to speech recognition, picture identification, text localisation, and others, machine learning encompasses examining computational algorithms that are dynamically improved through experience

(Charbuty and Abdulazeez, 2021). Machine learning makes it simple to tackle issues by creating algorithms that reflect a particular dataset. Machine learning is a branch of EDM that employs statistics to allow machines to learn and make responsive predictions based on data without needing to be taught to do so (Tolsgaard *et al.*, 2020). Machine learning is premised on automatically finding patterns in data using iterative and adaptive algorithms that constantly improve the machine's capacity to comprehend data (Sarker, 2021). As a result, the structure, attributes of data, and performance of the learning algorithms determine the efficacy and efficiency of a machine learning approach.

2.3 Contemporary developments of machine learning in several industries

Machine learning is an approach to AI. Within development businesses aiming to take a data-driven strategy to better their business by gleaning meaningful insights from the data they gather, ML has grown to be a prevalent subject (Injadat *et al.*, 2021). Companies can forecast changes in their industry and adjust their course of action using ML models. To enhance, explain data, and forecast consequences, ML use algorithms that iteratively learn from data. These algorithms let machines do specific jobs without human participation (Webb *et al.*, 2021). Without recognising it, machine learning provides numerous everyday advantages to end users in today's culture. For instance, thanks to machine learning, emails are trained to identify trends, so users do not need to wade through a substantial number of spam emails; most emails have already been filtered and sorted into the junk folder. Over 4IR growth, machine learning applications have emerged in many fields, including energy engineering, healthcare, and economics. Clinical care pathways, patient risk stratification for a particular disease, and diagnosis have benefited from machine learning predictions in the healthcare industry (Lopez-Bernal *et al.*, 2021). Machine learning algorithms in economics and finance have been used for portfolio creation, financial time series, and stock market analysis (Webb *et al.*, 2021). Machine learning models have been used to describe petroleum reserves, solar radiation, wind power, energy consumption forecasts, and reactor control optimisation in energy engineering and management (Lopez-Bernal *et al.*, 2021).

Similarly, Webb *et al.* (2021) state that machine learning is becoming increasingly prevalent in education and has been employed to enhance curricula, forecast learners' results, propose higher education programmes, and learn-modelling intelligent tutoring systems.

2.4 Machine learning in (programming) education

Technological breakthroughs in machine learning have the potential to improve or modify learning, and this prospect has consequences for what learners and teachers ought to comprehend regarding employing machine learning algorithms in the classroom. Presently, the South African education system is concerned chiefly with providing learners with material and hoping they will retain it (van Zyl, 2016). As a result, a learner's aptitude is determined by evaluating their capacity to remember previously taught information. However, this method is inadequate in developing learners' critical thinking skills, logical reasoning, and problem-solving abilities, which are crucial for programming (Peng *et al*., 2019). Moreover, Waite and Sentence (2021) argue that it does not address learners' diverse learning needs. As such, a gap in the teaching and learning of programming in South Africa needs to be addressed. Machine learning can play a crucial role in bridging this gap by enabling educators to provide personalised and adaptive learning experiences for learners.

Many schools are using technology to understand how machine learning may improve productivity and ease workloads (Tolsgaard *et al*., 2020). The following affordances of machine learning are noted to emerge over time:

- Tailored and personalised learning

Machine learning is versatile enough to accommodate the requirements of every learner, regardless of their pace of learning. Using these algorithms, it can figure out how the learner learns. Machine learning can signal to teachers that the learner has comprehensively understood the prior material (Nafea, 2018). This procedure guarantees that no learner is forgotten or disregarded behind. Even if they are the only learner in the class still having trouble understanding the material, this is still true. The machine learning method also allows educators to record each learner and assist them if necessary. This contrasts with the typical approach to education, which emphasises one-size-fits-all management and instructs each learner the same way.

- Content analytics

Machine learning algorithms can be used to evaluate teachers' knowledge in the classroom and assess if the content's quality is met (Nafea, 2018). This aspect refers to the machine learning algorithms where teachers look at content metrics to ensure that the teaching topics meet their learners' individual needs. The algorithms are also used to help determine whether the material presented to the learners is appropriate to each learner's intellect. Learners' learning progress and comprehension are improved since instruction is tailored for each learner.

- Optimising teaching and learning activities

Teachers clarify concepts that learners understand in traditional classrooms (Nafea, 2018). By advising teachers on the concepts, they should concentrate on promoting learning. Machine learning algorithms can be utilised to streamline these functions. Teachers will thus have more time to devote to more important activities, like ensuring that their learners thoroughly comprehend the subject content.

- Monitoring learners' progress

The teachers can keep track of each learner directly and evaluate their learning using machine learning algorithms (Nafea, 2018). Furthermore, these algorithms can provide learners with new learning methods, enabling teachers to discover the most effective strategies to educate their learners. Machine learning algorithms can customise the learning experience and resources to offer individualised learning with the assistance of adaptive tutors. They can also suggest resources or lessons to learners according to their characteristics, detect learners' grades or behaviour patterns, and communicate with them to give feedback.

South Africa's present approach to programming education is inadequate and creates an enormous gap in the learning abilities of critical thinking and problem-solving skills. By implementing machine learning techniques, such as supervised learning, in programming, educators can create personalised learning experiences for learners and promote higher-order thinking skills, leading to improved performance, and understanding of programming concepts.

2.5 Types of machine learning algorithms

Machine learning algorithms can classify profiles and patterns, create additional models and discoveries, and produce projections and suggestions tailored to the prerequisites of each learner (Luan and Tsai, 2021). Machine learning algorithms are classified into various categories based on the depending on the results of the algorithm. Machine learning is classed based on the data from which it can acquire information (Webb *et al.*, 2021):

1. Supervised learning - Whereby training data and accurate responses are provided.
2. Unsupervised learning - Whereby machines learn from a dataset by themselves.
3. Semi-supervised learning - Some data from the training dataset is absent, but the algorithms can still learn from it.

4. Reinforcement learning involves acting in a way that maximises benefit in a specific circumstance. Various applications and computers use it to determine the best potential action or course to take in each scenario (Luan and Tsai, 2021).

This study is concerned with applying the supervised learning algorithm to identify learners' programming challenges that impact their learning performance. It is a machine learning method in which the algorithm is supplied with labelled input data coupled with the output outcomes.

2.6 Supervised learning

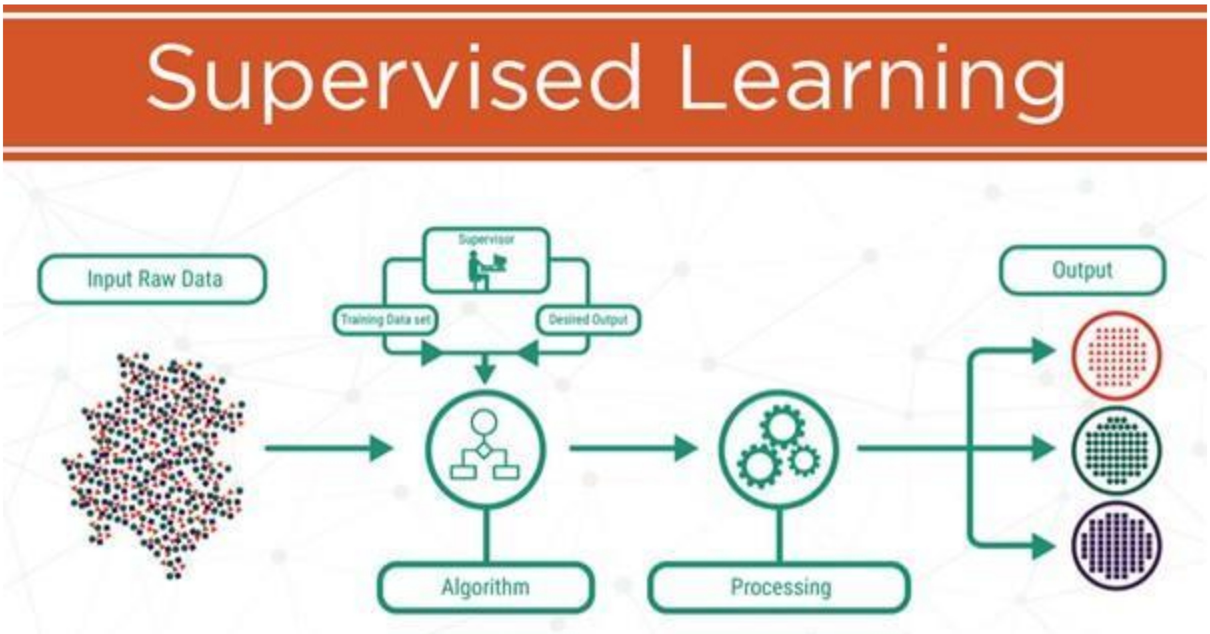


Figure 3: Supervised learning algorithms (Sarker, 2021)

Figure 3 represents the supervised learning techniques that postulate that programmes can detect patterns, detect errors, extract new data from input, and optimise their processing and output effectiveness and precision. Supervised learning is used when specific goals are established from an explicit set of inputs (Sarker, 2021). According to Sarker (2021), the idea of supervised learning in machine learning is to train a function that translates input to work using example input-output pairs. A supervised machine learning model divides the learning process into training and testing (Nasteski, 2018). The author points out that samples from the training data are used as input in the training stage, and the learning algorithm or learner learns the features and constructs the learning model. During the testing phase, the learning model predicts testing or production data using the execution engine (Sarker, 2021). In conclusion, the output of the learning model, tagged data, is the final prediction or classified data (Nasteski, 2018).

This is how supervised learning operates at the primary level: Consider, for instance, that an algorithm needs to determine whether a shopper would purchase bread this week. The algorithm will determine this by examining the last details and experiences, namely the information on the products the customer had previously purchased weekly. For example, if the customer routinely purchases bread, he is likely to do so this week. This example above shows the fundamental idea of supervised learning, which enables you to gather data or generate data output based on prior experience.

One research study aims to generate a model based on the supervised machine learning method that predicts programming obstacles for learners using previous results. To create models, supervised algorithms employ data from previously known classes. Their models then forecast the categories to which future unknown data were long (Osmanbegovic and Suljic, 2012).

2.6.1 Classification

Different supervised learning methods can be employed to produce a prediction model. Three significant prediction kinds exist classification, regression, and density estimation (Mueen *et al.*, 2016). However, in this study, the method for supervised learning used is classification. Classification is one of the primary methods used in supervised learning. The process of acquiring a mechanism that integrates the data into one of many predetermined classes is represented by different data classifiers (Mueen *et al.*, 2016). In a classification process, the model is created using samples that have already been classified to give a record a label or class. The training and testing stages of the classification approaches are classified into two parts. During training, a training set—a portion of the data that contains all the attributes and even the classes—is used to build the model. It is used to define a label or category for a new record if the class property is unknown after establishing a model (Mueen *et al.*, 2016). This means that each classification algorithm that uses induction learning is given an input data set that comprises units of data points and the classes to which they belong.

Building a model that enables the automatic classification of future data points based on a set of defined features is the aim of a classification approach (Osmanbegovic and Suljic, 2012). Classification algorithms may be compared to training a child to differentiate between a cat and a dog. To assist the child in learning to recognise cats and dogs based on their qualities, we may show him a variety of pictures of cats and dogs while indicating whether the animal is a cat or a dog for each image. Afterwards, the child could identify the animal's name when he sees a dog or a cat in the street. In this case, the input data's "label" is the animal's name that corresponds to it in the photographs (Lopez-Bernal *et al*., 2021). These systems receive a set of cases as input, each of which relates to a few classes and is described by the values it has for a predefined set of characteristics. They use a classifier whose output can precisely predict which type a new example belongs to. A model or classifier may be created using various methods, including Decision trees, Neural Networks, Naive Bayes, and Support Vector Machines (Khor, 2018).

The decision tree is a compelling supervised learning method (Mueen *et al*., 2016). The literature that provides reports on supervised learning algorithms suggests that the technique has a wide variety of classifiers (Osmanbegovic and Suljic, 2012), and it is impossible to determine the right one since they all have unique characteristics, notably different learning rates, training data requirements, classification speeds, and reliability. However, the decision-tree algorithm is used as a prediction technique under the classification model in this study.

2.6.2 Decision tree algorithms

A decision tree is a supervised classifier produced from a training set. A decision tree is a basic repetitive model for illustrating a sequential classification process in which a case is allocated to one of a unique set of classes based on a set of attributes (Khor, 2018). Nodes and leaves make up decision trees. Each node in the tree represents a test of a specific property, and each leaf represents a class (Mesarić and Šebalj, 2016). Typically, the test compares the value of an attribute to a constant. According to Khor (2018), leaf nodes provide a collection of classifications or a posterior distribution over all different classifiers that applies to all cases that reach the leaf. To classify an unknown instance, Mesarić and Šebalj (2016) note that the tree is rerouted down based on the results of the characteristics examined in consecutive nodes. When the model reaches a leaf, it is categorised based on its assigned class.

The decision tree algorithm is commonly used to find big or small data structures owing to its simplicity and ease of understanding (Khor, 2018). As a set of rules governs it, Khor (2018) denotes a tree-shaped structure representing sets of a decision, and the decision provides rules for dataset categorisation, which is a tree-shaped design representing sets of a decision (see Figure 4). The adaptation of the decision-tree algorithm is straightforward, and it is applied within the context of this study.

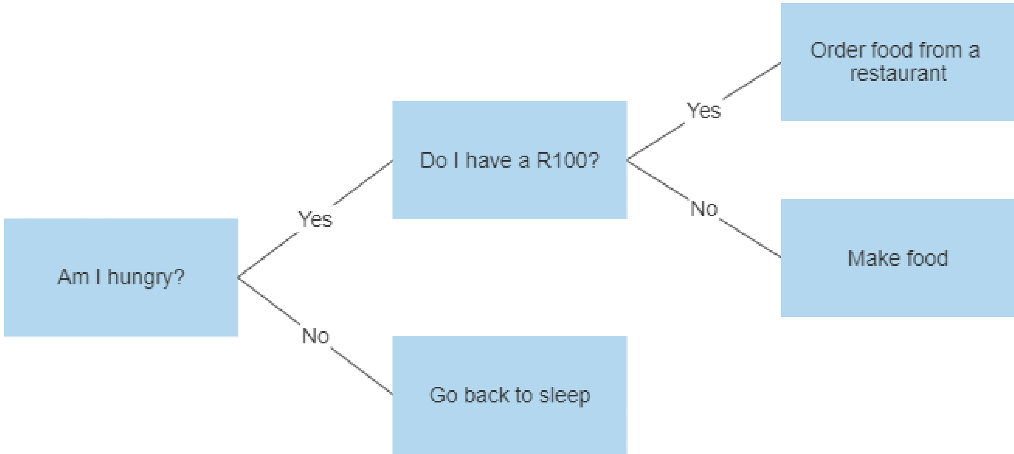


Figure 4: Decision tree implementation

A decision tree is a graphical representation of all the workable solutions to a decision. For instance, in Figure 4, a task outlined that when you are hungry, you should go to a restaurant or make food in the comforts of your home. You will create a decision tree, starting with the root node will first check whether you are hungry. If you are not hungry, then go back to sleep. If you are hungry and you have R100, then you will decide to order food from a restaurant. Moreover, if you are hungry and do not have R100, you will make food available in your house. In total, there are sixteen decision tree algorithms. Simple CART (classification and regression tree), NBtree, ID3, Reptree, and J48 are some algorithms. Among these, J48 is regarded as the most effective model-building method (Khan *et al*., 2021). The J48 algorithm does not need numeric attribute discretisation, in contrast to other algorithms. The discretisation is called discretisation when a numeric attribute's value is divided into fewer intervals. This procedure converts a numeric attribute's value into a nominal value (Khor, 2018). By utilizing the J48 algorithm in this study to identify programming concepts that teachers should focus on, this research contributes to filling the gap in the literature regarding the use of decision tree algorithms in machine learning to aid in the instruction of programming concepts.

2.7 Research gap

While there has been extensive research on machine learning algorithms in education, with most focusing on providing personalised learning (Nafea, 2018) and predicting student performance (Bhutto *et al*., 2020), there is a significant research gap in the specific application of machine learning algorithms to identify programming difficulties among learners. Studies conducted by Peng *et al*., (2019); Waite and Sentence (2021) note that programming syntax and concepts are complex and writing and executing programs requires learners to have logical thinking and problem-solving abilities. Even so, many learners struggle with programming, and it is critical to recognise these issues early on to prevent misunderstandings (Islam *et al*., 2021). Therefore, this study aims to fill the research gap by investigating the potential of a machine learning algorithm as a tool for identifying programming difficulties among learners. By doing so, this research contributes to the development of an evidence-based approach to programming education (Waite and Sentence, 2021) and provides insights into the potential use of machine learning algorithms in the South African education system.

2.8 Chapter summary

In this chapter, it was noted that machine learning algorithms have made routine procedures more accessible and efficient in the educational field. This has been a significant breaker in the education field. One of the most important advantages of its application, depicted in this study, is to support teachers in differentiating between issues that concern the entire class and those that subsequently affect individual learners. As a result, no learner gets neglected, thanks to machine learning (Nafea, 2018). The chapter has provided a comprehensive overview of several machine learning algorithms and their applications in education. By filling the research gap in this specific area, this study seeks to contribute to the existing body of knowledge on machine learning in education and provide insights into the potential use of these algorithms in the South African education system. The use of machine learning has nevertheless made the schooling system increasingly practical for teachers and learners. To optimise the learning experience using machine learning algorithms, two fundamental frameworks are essential in providing insight into using data which is explored in the next chapter.

Chapter 3: Theoretical and conceptual frameworks

"When you design a building, you start from a general philosophy, come down, and start from detail and come up. Only the theoretical architect believes that you can make the concept and then sometime, somebody will come to build it." – Renzo Piano.

3.1 Chapter overview

In this study, both theoretical and conceptual frameworks are used as guides for this study. According to Varpio *et al*. (2020), a theoretical framework is a logically created and interrelated collection of concepts and assumptions derived from one or more theories. On the other hand, a conceptual framework represents the state of available information, often through a literature review, and describes the methodological foundations of the research (Varpio *et al*., 2020). In this study, the theoretical framework involves the use of Educational Data Mining (EDM) to extract relevant information from databases, while the conceptual framework employs Learning Analytics (LA) techniques to comprehend the data and improve the learning process and environment. The use of both frameworks allows for a structured approach to achieving the objectives of the study, which involves modelling and implementing a supervised learning algorithm that identifies programming difficulties for learners and alerts teachers to intervene. Therefore, this chapter introduces the EDM and LA frameworks and how they contribute to the study's overall framework.

3.2 Theoretical framework

The field of data mining is concerned with discovering unique and helpful information from enormous volumes of data. In recent years, there has been a surge in interest in using data mining to study scientific topics in educational research, a field known as EDM. (Khalaf, 2016). EDM is a subset of data mining and machine learning that focuses on finding innovative ways to analyse educational data from a system (Pathan *et al*., 2014). It transforms raw data from an educational institution into patterns that may be used (Tan *et al*., 2016). EDM is a new area that seeks to create methods for studying the specific pieces of information accumulated in learning environments and leveraging those methods to effectively understand learners and the environments in which they learn (Mohamad and Tasir, 2013). EDM emphasises enhancing learner classifiers, particularly embodying their current knowledge, motivation, metacognition, and dispositions. The use of EDM to detect learner meta-cognition and self-regulatory competence within formal learning contexts could be helpful to scholars and researchers interested in learning to engage actively and steer their learning and meet the demands of achievement set within their courses (Berland *et al*., 2014). Patterns of learners' shifts throughout problem-solving tasks, for instance, can be identified to evaluate previously taught content and discern between more and less proficient learners based on their results. This brings forth the focus of this research on how EDM through supervised machine learning algorithms may help educators to assist them in identifying the challenges that learners face and enhance their academic performance.

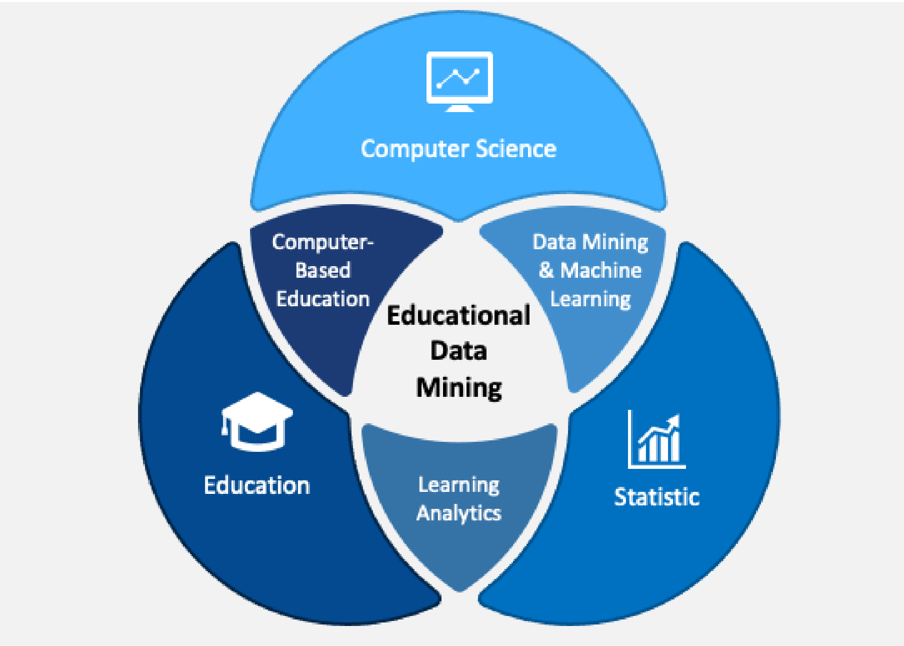


Figure 5: Principal areas involved in EDM (Khalaf, 2016)

One essential EDM uses the enhancement of learner models that predict learner traits or academic achievement in schools, universities, and other educational institutions (Khalaf, 2016). In all educational institutions, a substantial accurate prediction of learner performance is beneficial in various settings for detecting slow learners and differentiating between learners with poor grades and weak learners who are likely to have lower academic results (Ramaswami and Bhaskaran, 2009). Teachers, parents, and curriculum developers would benefit from the forecasting model's development by being able to advise learners on how to solve issues as well as tell them about how their present behaviour may be related to both positive and negative outcomes in the previous while they are studying (Khalaf, 2016). EDM permits the systematic, consistent, and accurate summary of learner behaviour, including examining how those behaviours intersect with other vital learning processes (Berland *et al.*, 2014). Learner behaviour may be monitored in real-time to see how it matures and evolves. EDM approaches have also been used to predict learners' readiness to learn new and varied materials from various courses, providing a tool for learning development. EDM approaches, particularly the supervised machine learning approach, link evaluations of different indicators of learner learning and learning processes to a multitude of other constructs, including connecting various indicators of learner learning and learning processes (Berland *et al*., 2014).

EDM examines construction and development processes, including the problem-solving and exploration areas where it is most often applied. As a result, EDM can be deployed to assess learner approaches and thinking processes, helping individuals grasp how learners develop strategies as they engage in learning activities. EDM and its methods have been used to research programming and develop programming abilities (Berland *et al*., 2013). They were given this background and that IT involves programming within the structure of the course, setting up EDM as the study's theoretical framework delineates the learners' programme creation practices and predicts if the learner is in jeopardy of failing to learn programming knowledge.

The vast data in databases makes estimating learners' performance more challenging. The descriptive information about data collection is successfully provided through descriptive statistical analysis (Tan *et al*., 2016). However, this is only sometimes sufficient. Estimated modelling approaches can notify educators and learners as soon as possible. To maximise success rates and effectively manage resources, it is beneficial to classify school learners according to their potential academic achievement (Pathan *et al*., 2014). As the amount of electronic data generated by schools grows, so does the need to extract useful information from these vast volumes of data (Tan *et al*., 2016). It is feasible to increase the quality of educational processes by employing machine learning algorithms on educational data.

Figure 6 shows the implementation of data mining in educational systems tailored to match every learner's needs. According to Romero *et al.* (2007), the learner should be provided with extracurricular activities, educational resources, and assignments that could benefit and enhance their learning. Moreover, the authors point out that the teachers may have the opportunity to provide feedback, arrange learners based on their need for guidance and help, identify past failures that learners encountered on concepts, and determine the most successful remedial measures. The directions of these remedial correctives increase learner performance; however, in this study, they are put in an algorithm to see the result of the growth (Romero *et al*., 2007).

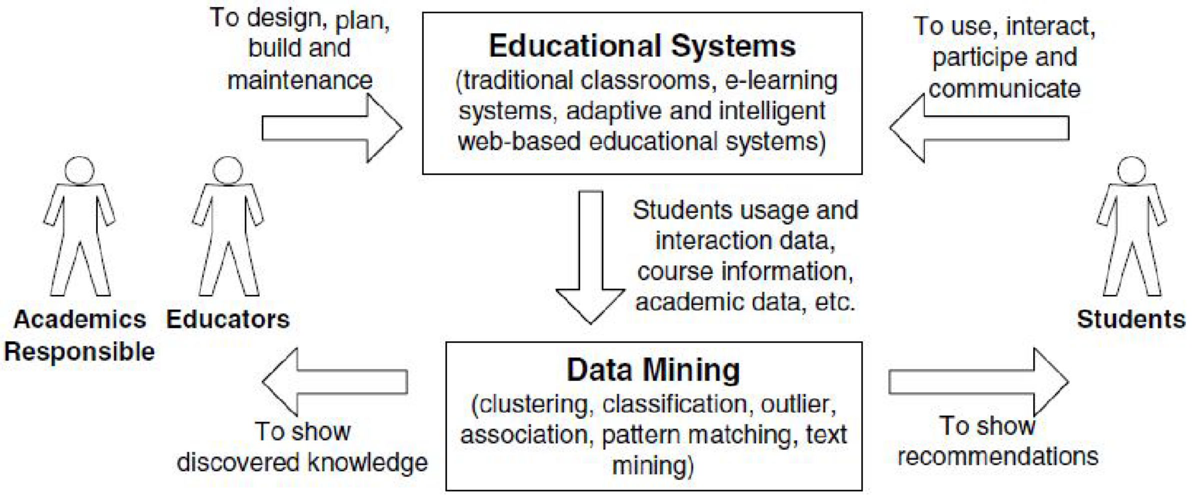


Figure 6: The cycle of applying data mining in educational systems (Romero *et al*, 2007)

3.3 Conceptual framework

EDM, as a theoretical framework in this study, is focused on designing, studying, and executing automated ways to find patterns in sets of data in the form of learner marks that would be challenging for teachers to analyse concepts that learners are struggling with because of the enormous amount of information within which they exist. LA is a conceptual framework to capture, measure, and report data on learners on concepts they are struggling with and their settings to comprehend and optimise programmed learning and the environments in which it occurs. LA is an emerging field of collecting large datasets to assess learning experience. The evaluation of the competencies of the respective programmes has sprung up because of the quick implementation of educational technology in modern classrooms and the deliberate efforts of educators to improve their formal and informal academic procedures (Kazanidis, 2021).

Steiner *et al*. (2014) state that LA may be executed for various stakeholders, each with its standards, demands, and objectives for the analytics process and its outputs. Teachers and learners are unquestionably the essential stakeholders in LA. As a result, the aims for employing LA are justified and align with the various viewpoints of its stakeholder groups. The following LA objectives are used in this study in conjunction with the stakeholders that the study is aimed at using the account of Steiner *et al*. (2014):

- Monitoring and analysis

This involves tracking and evaluating the learning experience, which teachers subsequently use as a foundation for enhancing their pedagogical techniques when teaching programming concepts.

- Prediction and intervention

To provide active interventions and assistance for learners who need help grasping programming concepts, it is essential to anticipate how learners might perform soon in terms of recognising early indicators for effective learning, failure, and prospective dropouts.

- Tutoring and mentoring

They enable teachers to advise while assisting learners with their whole learning process within the framework of programming learning activities.

When it comes to LA, a more contextualised and learner-centred viewpoint is necessary. For example, Kazanidis (2021) suggested a three-dimensional taxonomy and separated LA applications according to educational level:

1. Micro-level: Data is obtained by capturing a particular subject or in-class learning task.
2. Intermediate level: Data is acquired by capturing complete teaching and learning or unit.
3. Macro-level: Data is collected by documenting a collection of subjects or instructional programmes.

From that frame of reference, this adoption of LA as a conceptual framework is at the micro and the intermediate level, as the data that the study draws is based on using past IT results of learners to inform teachers about programming concepts; they need to pay attention to. The data extracted at the micro and intermediate levels are intensive. Intensive data is a deep but restrictive set of data with few participants but detailed assessments of various variables (Steiner *et al*., 2014). Therefore, the methods used to extract the data (learners' results) into critical patterns must be considered. LA is divided into five primary classes: prediction methods, structure discovery, relationship mining, discovery with models, and distillation of data for human judgment. Steiner *et al*. (2014) stress that the methods used are determined by the objectives of the analytics activities, in addition to the type of data gathered; hence, this study considers the use of prediction algorithms and data distillation for human judgement.

- Prediction

The primary objective of this approach, which is the most often used in EDM, is to build a model to predict or infer a specific variable (such as a mark or performance score) from a collection of other indicators in the educational dataset (Steiner *et al.*, 2014). It was determined in the literature review section that classification would be the study's method of prediction. This procedure is made more accessible by categorising training and testing datasets.

- Data distillation for human judgement

Steiner *et al.* (2014) note that data distillation for human judgement is a conservatively common approach in LA. Still, it is not regarded as an EDM method because it gives teachers rapid access to reports and visualisation tools of learner data for their interpretation judgement. It also assists decision-making and pedagogical action. This approach in this study provides teachers with a description of programming concepts they need to use as a visual tool in the shape of a decision tree to flag these concepts. With immediate access to the tool, it can allow teachers to gain knowledge on concepts their learners are struggling with, thus, allowing them to intervene at an earlier stage.

Both approaches necessitate descriptive and regulatory analytics as a classification method within the framework. Descriptive analytics focuses on what has already occurred. It seeks to find patterns by integrating records of learners, whilst regulatory analytics focuses on the demands that matter and the variables that impact how well learners learn and offers suggestions for upcoming events (Kazanidis, 2021).

3.4 Chapter summary

EDM and LA are two developing disciplines with many commonalities regardless of variances in their backgrounds and applications. In this study, they are integrated as a theoretical and conceptual framework. As a conceptual and theoretical framework, the combination of EDM and LA has a strong potential to influence the current education models (Tan *et al.*, 2016). This would provide fresh insights into learners' learning routes and enhance learning strategies using data. LA and EDM are used in this research to understand learning processes better and enhance learning through educational systems. They are employed in this study to educate and assist learners, teachers, and their schools, helping them to comprehend how these potent tools may use data sources in the form of prior IT results that lead to significant gains in learning and educational performance.

Chapter 4: Research methodology

"What is research but a blind date with knowledge?" - Will Harvey.

4.1. Chapter overview

In the previous chapter, there was coverage of the comprehensive theoretical and conceptual frameworks which have been selected for this study and emphasises that both EDM and LA were employed to develop a comprehensive analysis of preceding results to determine topics that learners are experiencing difficulties with and ought to be raised to the teacher's attention using a supervised learning algorithm. The purpose of this chapter is to show the methodological approach employed and the path taken to fulfil the study's purpose, which is to inform IT teachers about the topics or concepts that learners are struggling with and allow them to intervene at their earliest convenience. According to Merriam (2009), when conducting research, one should be as thorough as one can be in their search and study of the topic. This chapter explains the methods and approaches used to gather, arrange, and evaluate the qualitative data obtained through semi-structured interviews with IT teachers. In addition, this chapter outlines the justification for the participants chosen for the study, along with the methods used for data analysis. Lastly, the study's ethical considerations are also addressed in consent to the reliability and validity of the study.

4.2. Research design

All the components of a research study are kept together by the research design, which would be alluded to as the research structure. Numerous people have described research design uniquely, but a unique one is captured by Akhtar (2016, p.68) “A research design is the arrangement of conditions for the collection and analysis of data in a manner that aims to combine relevance to the research purpose. Research design is the plan, structure, strategy, and investigation concaved to ensure search questions and control variance.” Therefore, the research design entails laying out the whole strategy for the research.

This study follows a qualitative case study research design of schools with IT as a subject around Ekurhuleni South. A qualitative case study is a research design that helps examine a phenomenon within a specific context using numerous datasets. It does so by using a variety of perspectives to disclose various facets of the event (Rashid *et al.*, 2019). Qualitative case studies enable the researcher to obtain a comprehensive picture of the research study and make characterising, comprehending, and describing a research problem or scenario efficient

(Baškarada *et al*., 2013). This leads to the study using a descriptive case study on the programming concepts in which Grade 10 learners in IT are struggling and calling for the teacher's attention. This approach is an exploratory type of inquiry that allows for extensive engagement with participants in the study, culminating in a detailed image of the study unit (Bloomberg and Volpe, 2019). A proponent technique to describe the case within the context of this study was descriptive rather than exploratory. Yin (2018), as cited in (Bloomberg and Volpe, 2019), describes an exploratory case study as used to examine problems where there would be no single set of outcomes for implementation. A descriptive case study represents an actual situation whereby an intervention in the fundamental phenomena is provided based on the study's findings (Bloomberg and Volpe, 2019). The study's natural spectacle is programming concepts teachers need to pay attention to when teaching based on previous results.

Since this study follows a descriptive case study, the inductive process was a conducive component. According to Merriam (2009), in positivist research, researchers collect information to construct conceptions, hypotheses, or theories instead of deductively testing ideas. This means that approaches are based on my observations and instinctive understandings while in the fieldwork. During the fieldwork, sources of evidence gathered from interviews and documents of the learners' marks are integrated and organised into broader themes. These themes form a preliminary hypothesis regarding the challenges teachers need to pay attention to inductively, which are developed from the data in the qualitative inquiry. The ontological and epistemological underpinnings are provided below to discuss the paradigm disposition in this study.

4.2.1. Research paradigm

Ontology is the study of existence, and ontological premises deal with the nature of reality. Using ontological and epistemological presuppositions, research paradigms are established. The research methodology used in the study is directly influenced by the ontological and epistemological perspectives held by the researchers. A researcher's ontological perspective might be characterised as realism, for instance, if they think that social reality is outside of an individual's mind and exists beyond the person asking about it (Cukurova, 2018). While ontology is interested in knowledge, epistemology deals with how humans obtain special knowledge about the entities in question. Epistemological assumptions are interrelated to considerations about the nature of knowledge, how it is generated, what shapes it takes, and how it could be transmitted (Cukurova, 2018).

There are several paradigms in educational research, including positivism, interpretivism, Marxism/critical theory, postmodernism, and critical realism. However, the research paradigm applicable to this study is critical realism. The epistemic and ontological views of this study have been discussed in greater detail above, and it is apparent that the essential philosophical principles of this study are drawn from the critical realism paradigm.

Critical realism paradigm

A viewpoint in the philosophy of science known as "critical realism" is frequently associated with Roy Bhaskar. The essence of critical realism is that there is an existence of an external reality apart from the understanding of it (Bhaskar and Hartwig, 2016). In this regard, individuals can be identified as objects as they exhibit traits and skills that help people form their beliefs and knowledge, which is the reflexive element of reality. Since humans are flawed and tend to display things specific to a particular area, period, and culture, this makes existence complicated, according to Sayer (2000). This indicates that an individual's world representation is circumscribed and controlled. Consequently, the critical realism paradigm's application in this study's ontological stance must be seen in two ontological views from the theory: (1) Critical Realism as a theory that can govern machine learning algorithms in educational settings and (2) how the output of the algorithms leads IT teachers to focus on programming concepts they can use to encourage that can foster creativity for problem-solving programming tasks.

The first application of this paradigm entails developing causal mechanisms and the context in a critical realism manner, drawing on action research and classification approaches. According to Fox and Do (2013), a critical realist determines the causal context and process that can allow an action to result in an inevitable consequence. This frame of reference can be drawn into the purpose of this study which is to use past IT results to inform teachers about concepts they need to pay attention to using a supervised learning method. According to the critical realism viewpoint, the causal mechanism and context cannot be known initially (Fox and Do, 2013). In this study, abductive reasoning is frequently employed to do. This entails extrapolating causal mechanisms and context for a particular event through iterative cycles of comparison to theories and findings. Incoming data must be organised to facilitate analysis and presented in comparison with notional models of operational state circumstances to permit automated assertion (Fox and Do, 2013). In this study, the incoming data are learners' marks that are organised and classified against programming concepts. Relational methods are used to extract this data through the process of classification using data mining methods (Fox and Do, 2013).

Hence, the adaptation of EDM and LA in this study. The analysis of these assertions is governed by supervised learning. This leads to the output being represented through a decision-tree algorithm.

The second application of the critical realist in this study is studying the nature of the causal mechanisms of teaching and learning environments in which programming lessons take place. Following Potter and Lopez (2001), humans and structures are knowledge objects. This implies that in education, learners, teachers, and social structures function as catalysts for the surroundings, procedures, and results of learning. Teaching and learning take place in an open system since the classroom is a structure that is based on the real world. Consequently, before reaching out to the learner's marks as the conclusive aspects of the teacher's efforts, the adaptation of critical realism is to find the causal mechanisms in the form of concepts that need the teachers' attention the most whilst also looking at strategies they used to teach those concepts. The critical realist account's core tenet is the causal mechanism because it suggests a natural, stratified world with generative processes that produce events that can be objectively observed (Bhaskar and Hartwig, 2016). Objects and practices are examples of mechanisms and their attributes that exist independently of the knower and have a sporadic impact on people and events (Potter and Lopez, 2011). According to Sayer (2004), understanding the characteristics of the thing or structure that has the power helps to explain why a specific mechanism occurs. This is done to comprehend the driving force behind an emerging occurrence. The causal mechanism's utilisation in connection with this research is to outline the programming concepts they need to pay attention to. This can assist teachers in looking at causal mechanisms in their teaching strategies and other structural factors that inhibit the learners' understanding of these programming concepts leading to them performing poorly in these topics during summative assessments. The research approach depends heavily on one's ontological and epistemological viewpoints. The study's next section examines the study approach used.

4.2.2. Research approach

This study belongs to the qualitative approach, which is one of three options for doing research, along with quantitative and mixed methods. To differentiate between these research approaches, one must look at each course. Qualitative research is concerned with understanding aspects of social life, whilst quantitative analysis is concerned with measuring something (Merriam, 2009). Mixed methods integrate the latter research approaches in one study to obtain knowledge about the subject (Bloomberg and Volpe, 2009). Neither quantitative nor mixed methods are used in this study. However, qualitative research is applied in this study due to its flexibility in answering the “what.” The issue of "what" is addressed in qualitative research (Merriam, 2009). It knows what necessitates comprehension of the subject matter and its numerous components (Bloomberg and Volpe, 2009).

Qualitative research is “an umbrella term covering an array of interpretive techniques which seek to describe, decode, translate, and otherwise come to terms with the meaning, not the frequency, of certain more or less naturally occurring phenomena in the social world” (Merriam, 2009, p.13). This means that at the heart of qualitative research is a thorough comprehension of a subject matter in all its actual complexity and the capacity to define, explain, and express that knowledge. This background gave the participants in this research the freedom to respond openly using their own words as they described in detail the features of the ways that their learners learn programming. Additional inquiries were made where clarification or explanation was required. Therefore, applying a qualitative research methodology is well suited to fostering a thorough understanding of a social situation as seen through the eyes of the study participants. The research methodology for this study is covered in the section that follows.

4.2.3. Research methodology

The process of conducting research is known as the research methodology. According to Scott and Morrison (2006), this contains the steps the researcher takes to answer the research issue and the justification for employing specific procedures and methods to plan, collect, and generate diverse information on educational processes. A research methodology is therefore regarded as an approach that governs how researchers perceive the topic being explored in research areas and the justification for choosing strategies or methods over others to handle the defined research problems. Grounded theory, ethnography, action research, phenomenological research, and narrative research are qualitative research methodologies (McNiff, 2013). Action research has been selected as this study's methodology because it is a method of practice that entails acquiring data, reflecting on the action as it is shown in the data, producing evidence from the data, and asserting knowledge based on inferences made from validated evidence (McNiff, 2013).

4.2.4. Qualitative data collection: Research methods

Questionnaires, interviews, observations, and case studies are research methods used by researchers to collect data on the topic under investigation (Merriam, 2009). The research methods mentioned in the preceding sentence are used to gather qualitative data. Individual semi-structured interviews were used in this research to accomplish the objectives of the study and endeavour to formulate a solution to the proposed research.

Interviews

Across all qualitative research approaches, part, if not all, of the data is gathered through interviews (Bloomberg and Volpe, 2019). Moreover, the same is true for this research. An interview is a procedure whereby a researcher and a respondent interact in a discussion concerning identified research topics. The person-to-person interaction, in which one person obtains information from another, is the most prevalent type of interview used in this study. Person-to-person interviews are conversations with a goal (Merriam, 2009). The primary purpose of an interview is to get specific information. According to Merriam (2009, p.88), "Interviewing is necessary when we cannot observe behaviour, feelings, or how people interpret the world around them. It is also necessary to interview when we are interested in past events that are impossible to replicate." In the context of this study, the latter in the quotation, using past results in the subject of IT, there is considerable interest in finding concepts that teachers need to pay attention to that learners seem to find challenging that cannot be directly observed as assessment is usually done late whilst also interested in finding out about how teachers shape learning experiences mainly when they teach programming concepts. Therefore, the interviewed population persons were four teachers to gain their perspective. Moreover, interviewing is also the optimal strategy to use when doing a detailed case study of a few samples of individuals (Merriam, 2009).

Interviews with each participant in a semi-structured manner

Different sorts of interviews can be classified in several ways, including highly structured, semi-structured, and unstructured/informal formats. In the context of this study, interviews were conducted in a semi-structured format. The characteristics of this interview format are outlined by Merriam (2009), and this characterisation allowed the researcher to obtain the necessary data:

- The interview process offered a combination of structured and unstructured interview questions.
- All questions were used with flexibility.
- Precise information is expected from all participants – Including diagnostic assessments of past results.
- Most of the interview was directed by a series of questions or problems to be addressed.

- There were no fixed terms or arrangements in which the questions were asked.

This set of characterisations allowed the researcher to respond using the participant's evolving outlook and fresh insights. Using semi-structured one-on-one interviews created an environment which allowed engagement in open-ended discussions with the teachers, who were then able to freely recite the techniques they use to teach programming and provide a comprehensive description of the intervention tools they wish to have on the occasion that learners do not understand (see Appendix E). The concept underpinning qualitative research is that rich data in a social context can only be gathered through an interactive process between the researcher and the participants (Bloomberg and Volpe, 2019). There was an attempt by the researcher to clarify the meaning of the findings from the viewpoint of the study participants and to do so. Data is collected directly from participants.

As the researcher, I was the critical instrument for the collection and analysis of data. Since the primary goals are description, comprehension, interpretation, and communication (Bloomberg and Volpe, 2019) from the interviews. The interviews were taped to help the researcher correctly transcribe what the respondents were saying because the audio tapes were listened to characterise the data during data analysis iteratively. The researcher had to be aware of his assumptions when transcribing the data to avoid bringing his perspective of what the respondents had said, which would weaken the study’s validity. Demonstrating that the data is trustworthy and credible improves the study's integrity; this is covered in more detail later in the chapter. Moreover, Bloomberg and Volpe (2019), essential to the qualitative approach, is centring on how participants in the study think regarding a particular phenomenon. The duration of each interview is shown in Table 4 below.

Table 4: Duration of each semi-structured interview

Participant’s name[1]	**Duration of each interview**
Teacher A	41 minutes and 30 seconds
Teacher B	19 minutes and 28 seconds
Teacher C	29 minutes and 36 seconds
Teacher D	35 minutes and 55 seconds

[1] Pseudonyms were used in this study to protect the identities of the teachers

Given the format of semi-structured interviews, it is apparent from the table above that the time needed to complete each interview varied. Some teachers had significantly more information to disclose than others. This affects the length of discussions as they differ from one another (Merriam, 2009). Even though each interview had a 30-minute time limit, due to the open-ended structure of semi-structured interviews between the participant and the researcher, the researcher was unable to discourage some participants from detailing their views on teaching programming concepts. When the interviews were conducted, it was negotiated that they are completed after school to avoid interrupting the teachers' lesson schedules. However, the first half of the interviews were conducted in the morning because the teacher had free periods. In comparison, the second half of the interviews were completed after school. Each of the interviews occurred on separate days.

The collection of the IT learners' marks

Due to the nature of the study being descriptive, additional data in the format of documents which previous marks were requested to support the research findings. Teachers were informed ahead of the interview to prepare these documents. Before and during the interview, it was reiterated that these marks would be inserted into the decision-tree algorithm to flag topics that learners have difficulty with about programming concepts in IT.

4.3. Research sampling approach

Several scholars have suggested that choosing research samples for a qualitative method is premised on the assumption that the respondents have some background understanding of the subject. Different sampling strategies can be found within two sampling techniques: probability sampling and non-probability. A probability sampling procedure is often used in quantitative studies to generalise results (Yin, 2018). This procedure includes random, systematic, stratified, and multi-cluster stage sampling (Bell and Bryman, 2007). In contrast to the probability sampling procedure, a non-probability sampling procedure can select participants from a particular region to participate in the study (Babbie and Mouton, 2007). According to Bloomberg and Volpe (2009), non-probability sampling techniques include purposive, convenience, and dimensional sampling. Since this action research aims to provide accounts of specific populations situationally instead of generalisations, a convenient and purposive sampling method is used in this study. Purposive sampling is a method of discovering and picking instances that will efficiently use limited research means by choosing the respondents who are most likely to provide relevant and valuable information (Campbell *et al.*, 2020). The rationale of purposeful sampling is to select the sample from a wealth of information to provide insight and knowledge of the studied topic (Bloomberg and Volpe, 2019). With convenience sampling, participants are chosen based on their immediate availability. This accessibility is often measured in terms of the closeness of the participants' location to one another (Campbell *et al.*, 2020). Therefore, the rationale for using the convenience and purposive sampling techniques is based on the idea that, given the study's objectives and goals, particular kinds of individuals may possess distinct and significant viewpoints on the subject matters related to the research question, and hence must be included in the sample.

4.3.1. Sampling strategy

The non-probability sampling approach is selected in this action study, focusing on purposive and convenience sampling techniques. The purposive sampling technique was used to determine the participants of the study. This technique enabled the selection of specific participants based on their characteristics concerning the study's rationale. IT teachers with more than two years of teaching the subject were selected since they would have pertinent knowledge of this research study. In addition to purposive sampling, convenience sampling was used to determine the study's location. The selected schools are based in Ekurhuleni South. Given the proximity of the schools, it was convenient to travel around those schools on the scheduled interview dates.

4.3.2. Participants

Four IT educators for Grade 10 were the study's participants. The factors listed in Table 5 below guided the selection of the educators.

Table 5: Teacher selection criteria for the study

Teaching subject	IT
Experience	The teacher needs over two years of teaching IT, Grade 10.
Province	Gauteng
Area	Ekurhuleni South

The four teachers who took part in the study fulfilled the requirements for action research, which, according to McNiff (2013), the sample size for action studies should typically not surpass ten, and it would consist of four or five members. The choice to have four teachers was made to optimise the variety of available information for the data analysis process and to have acceptable data.

4.4. Data organisation and analysis

The procedure by which the researcher bases principles, organisation, and relevance to the data they have acquired are known as data organisation and analysis (Wong, 2008). Several scholars claim that the processes of gathering and analysing data are intertwined in qualitative research (Wong, 2008; Merriam, 2009; and McNiff, 2013). Merriam (2009) asserts that there is no specific point at which data analysis starts. In other words, data analysis involves giving initial observations made during data gathering and final compilations significance. The data analysis began after the interviews to find patterns within the participants' responses. This means that the study followed a thematic analysis. A qualitative analysis technique known as thematic analysis examines classifications and presents themes (patterns) correlated to the data (Merriam, 2009). The six phases of thematic analysis recommended by Braun and Clarke (2012) were used to find commonalities across various data types that could be used to conceptualise and interpret the data. These phases are familiarising oneself with the data, generating categories, generating themes, reviewing themes, defining, and naming themes, and producing the report (Braun and Clarke, 2012). Whereas the phases were presented sequentially by Braun and Clarke (2012), the process in this study was not linear; specific repetitive steps during analysis contributed time to the procedure.

Familiarising oneself with the data

This phase, universal to all types of qualitative analysis, is fully immersed in the data by reading and rereading textual information (such as interview transcripts) and listening to audio recordings (Braun and Clarke, 2012). This phase entailed reading and rereading the interview transcripts, making notes in the margins of the transcript documents on the teachers' choice of words, and afterwards, writing a summary for each interview transcript following the research objectives. After that, the interviews with the research participants were tabulated during the transcribing process using Word Processor (see Appendix A). The current study's method is action research. Therefore, it involved carefully examining how Grade 10 IT educators shape learning experiences in programming concepts, which made it possible to develop the coding-based systematic analysis of the data.

Generating categories

Categories are the foundations for recognising the meaning of patterns in data, and a fundamental organising principle supports them. According to Braun and Clarke (2012), this phase entails selecting data units that possess the capacity to provide information pertinent to the research questions. The analytical procedure entailed finding categories that would make sense to the data. The data extracts were created so that the categories emerged from the transcribed data. This inductive strategy is favoured to provide desired responses to the research questions.

Given that this is a descriptive case study, the inductive coding approach encouraged the researcher to first focus on what the teachers were saying, along with the descriptive observation notes, to find the core ideas and assumptions (Merriam, 2009). It is critical to examine appropriate category names and provide extensive summaries to describe the categories throughout the creation of the category list to ensure that the implications connected with each category seem evident. The category names used to describe the data qualitatively were derived from the actual words and behaviours represented in the data (Braun and Clarke, 2012). The descriptive categories from the interviews with the participants created during the transcribing process are shown in the table below:

Table 6: Generating initial categories from the interviews

Initial categories
The utilisation of the CAPS document to structure objectives.
Integration of a planning process.
The inclusion of real-life examples.
Collaborative learning
Solution development
Guidance by the teacher
Measuring learner experiences

Generating themes

The analysis emerges as the attention turns from categories to themes in this phase. According to Braun and Clarke (2012), a theme is a significant aspect of the data related to the research question and signifies a level of patterning in the participants' responses. Throughout this phase, the categorised data were reviewed to find instances of commonality and duplication between categories. The rationale behind this is that the direct approach to creating main themes and subthemes, which are the components of a theme, is compressing or grouping categories that share a common trait to capture and describe the substantial and consistent pattern in the data (Braun and Clarke, 2012).

Reviewing potential themes

Quality assurance is the focus of this phase. The first process is to evaluate the themes of the generated data extracts and determine whether the theme made sense in the data context. In this phase, the emerging themes were assessed in connection to the categorised data and the overall dataset through a sequence (Braun and Clarke, 2012). On the categories that did not function, a few were eliminated or transferred to a different theme; instead, the theme's confines were redrawn to capture the relevant data more accurately.

Defining and naming themes

When outlining their themes, researchers must be able to describe succinctly what makes each one special and distinct. An effective way to check this is to summarise each theme's main points in a few phrases (Braun and Clarke, 2012).

Table 7: Main themes

Main themes
Theme 1: The role of problem-based learning in programming
Theme 2: Automated feedback tool

Producing the report

The analysed data is presented in this phase to implement the findings (Braun and Clarke, 2012). The information supported by extracts, organisational tables, and a data map were interwoven in this final analysis stage to implement the findings. This is accentuated by the

WEKA tool concerning the chosen conceptual and theoretical frameworks. This detailed presentation is in the next chapter.

4.4.1. Using the WEKA tool for data analysis concerning the theoretical and conceptual frameworks

WEKA is open-source and free software containing various machine learning algorithms. According to Khalaf *et al*. (2016), WEKA is a set of innovative machine-learning algorithms and data preparation utilities. It is meant to allow users to test current methods on new datasets in various ways swiftly. It offers comprehensive assistance for the entire experimental data analysis, including data preparation, statistical measures of learning schemes, and visualisation of the input data and learning results (Khalaf *et al*., 2016). According to Salihoun (2020), it is an available tool to model and identify correlations and processes in educational data. This indicates that it provides resources for EDM and LA tasks like clustering, regression, classification, association rules mining, and visualisation (Khor, 2018). Considering the applicability of classification as a method for supervised learning in this study, a classification model is being constructed using training data. Each data point in the dataset is pre-classified (Khor, 2018), meaning its class label must be known; second, the model created in the previous phase is evaluated by giving class labels to data points in a test dataset (Salihoun, 2020). As established by the class label attribute, each tuple or sample is taken to belong to a predetermined class. Decision trees are used to visualise the model. In simpler terms, the tool is used as an EDM and LA activity in this study to analyse and model datasets and verify the output models once attributes have been created, then to verify that outcome variables have been labelled, and eventually to guarantee that data has been collected and suitably structured (Salihoun, 2020).

Consequently, this study uses the supervised learning algorithm's analysis to flag programming concepts that learners are struggling with, thus requiring a teacher's assistance. Therefore, the WEKA toolset provides extensive pre-processing tools and learning techniques. Given the background of the WEKA tool in building the decision-tree algorithm, there are three steps which were essential in the in-depth analysis: data pre-processing, the exportation of data, and generating the data output. The processes of each step are outlined in the next chapter.

4.5. Ethical considerations

In any research study, it is essential to protect the safety of participants by adhering to the proper ethics rules. There in-depth nature of the research process in a qualitative approach lends ethical considerations a special significance (Arifin, 2018). When conducting a face-to-face interview, ethical implications are more apparent. According to Babbie and Mouton (2007), an argument can arise when a researcher interacts with an issue that is likely to result in conflict, hence why a researcher must always be mindful of the ethical consideration of the study. Thus, the researcher must refrain from doing anything that could be considered unethical a could infringe upon the rights of others. This section provides a guideline on ethical considerations followed throughout the study.

4.5.1. Access to the schools

Initially, I sought ethical approval from the University of the Witwatersrand's Ethics Committee. Simultaneously, the GDE (Gauteng Department of Education) permitted me to study at the selected schools because it is part of a comprehensive research report after applying for permission to access schools (see Appendix B). Before the investigation could begin, the research proposal of this study was approved at this time (see Appendix C), along with the GDE and Wits Ethics Committee acceptance and clearance letters to gain entry to the schools and conduct interviews with the teachers. This insinuated the genesis of the methodology part of this research, where a letter was written to the school principals requesting permission to use their schools as sites for interviews (see Appendix D). This is because participants must choose a setting acceptable for them and feel confident enough to disclose any information to express themselves frankly and openly (Babbie and Mouton, 2007). From that frame of reference, it was constituted that teachers would feel more comfortable responding to questions openly if they were in a familiar environment, such as their classroom. Therefore, they could withdraw their agreement to participate in the study at any moment without worrying about being casual or even without me making a case for them to continue to take part.

4.5.2. Privacy protection for schools and their teachers and learners

Preserving the participant's identification is a concern that is at once related to protecting their rights and welfare (Arifin, 2018). Since data was collected on how they create learning experiences while teaching programming concepts, the marks of their learners and the identities of all participants should be preserved. Arifin (2018) supplied the guidelines for anonymity and confidentiality to guarantee that research participants' identities are protected. The interviews were conducted after school, so the teachers were assured that their standard classroom practices would not be harmed. Before the interview, teachers were also told they might withdraw from the project anytime. Exacerbated by the interview, teachers were asked to share their learners' earlier examination marks so they could be tested in a supervised learning algorithm. Despite gaining access to the patterns of the learners, the teachers disclosed their marks by following the principle of "Anonymization” in the POPIA (Protection of Personal Information Act) Act of 2013 (SAGG, 2013). Personal information is transformed so that a data subject can no longer be directly identified. In short, the names of the learners, schools where the investigations were conducted, and teachers are not shown in this research.

4.5.3. Explicit consent

Two types of ethical methods for educational research inform consent: covert and overt. When doing their work covertly, researchers keep the nature and purpose of their study from participants (Scott and Morrison, 2006). This study breaches the ethical premise of explicit consent (Bell and Bryman, 2007). Covert research suggests that people cannot decide whether to participate in the study. Nevertheless, governed by the informed consent premise, the research could be investigated overtly. Instead of using the covert method, this research adopted an overt approach to ethical research. According to Scott and Morrison (2006), the procedures used in an overt research ethical method entail the researcher openly revealing to study participants the intent of the research, the rationale for which it is being conducted, and the way they will use the information participants provide. The participation information sheet and consent letter outlining the purpose of the study were sent to the principals and teachers of the IT subject. The participation information sheet and the consent form were read before the interviews, and both documents were approved. Accordingly, when the four teachers were interviewed, it was clear that this study was for my Master of Education Research Report. However, the data they supplied could also be used for journal papers published in academic journals and conference proceedings to further the discipline of leveraging machine learning algorithms in education.

4.5.4. Permission of the participants to withdraw from the study

The right of the participants to opt-out of participating in the current study is another approach to ethics that was upheld throughout the study. Potential research participants should have all the information they may need to decide whether they would like to participate in the study (Bell and Bryman, 2007). Each participant was informed of their choice to decline participation in the study before the interviews began, reiterating that their involvement in the research was optional. This implied they had the right to revoke their consent to take part at any moment and for any reason (Babbie and Mouton, 2007). Nevertheless, during the research, neither of the chosen participants withdrew their participation.

4.6. Assuring the integrity of the research outcome

Several elements are considered crucial that enhance the study's reliability. For the study to be recognised in the canon of knowledge and to be considered suitable for application in a range of ways and methods, Bell and Bryman (2007) contend that the research study must adhere to the criteria of ensuring that it is exact, dependable, and generalisable. The measure that was essential in guaranteeing a study's trustworthiness includes authenticity, conformability, credibility, dependability, and transferability (Lincoln and Guba, 1985, cited in Elo *et al*., 2014).

4.6.1. Authenticity

The degree to which researchers correctly and accurately present a multitude of realities is called authenticity (Elo *et al*., 2014). To ensure that the study is authentic, the principles outlined by Bell and Bryman (2007) were followed:

- Fairness: The research stands for the perspectives of IT teachers without asserting the researcher's biases.
- Ontological authenticity: This study arrives at a better social milieu by asserting the perspectives of IT teachers on the need for intervention tools that can be used to flag concepts that learners are facing challenges.
- Educative authenticity: This study appreciates different perspectives in answering the study's questions.
- Catalytic authenticity: Concurrent with ontological authenticity, the study endorses machine learning to advance and improve learning experiences in educational settings.

4.6.2. Conformability

Bell and Bryman (2007) note that the extent to which the discoveries result from the inquiry's emphasis instead of the researcher's inclinations is alluded to as confirmability. I listened to the audio recordings several times to verify that the transcripts accurately reflected the data provided by the participants and kept a reflective journal that described all the events that occurred in the study environment to ensure that there was an act of honest conduct, or that no core views or theoretical proclivities were appearing to influence the conduct of the research and conclusions derived from this.

4.6.3. Credibility

The researcher must confirm that the research findings fairly reflect the data gathered from the participants in any empirical research. Establishing the credibility of conclusions requires both making sure that studies are conducted following standard practices and presenting study results to the individuals in the social environment that have been examined for confirmation that the researcher has correctly understood that social reality (Bell and Bryman, 2007). Distinct complementing evidence-based sources were used to achieve credibility in this study and get differing views on the research topic. Data in this study were collected in two forms: interviews with multiple IT teachers and the records of the learners' marks to be inserted into an algorithm. The numerous reference points allowed the study's findings to be pertinent to the research aims.

4.6.4. Dependability

Dependability focuses on determining the importance of research compared to the standard of credibility. According to Bell and Bryman (2007), this calls for the researcher to adopt an auditing strategy by making sure that thorough records are kept of every step of the research process, including problem formulation, sampling procedure, research notes, transcriptions, data analysis choices, and so forth—in an approachable way. The supervisor served as the auditor for this research to address concerns with dependability. In his role as an auditor, my supervisor went over the data that was gathered, the interview transcripts and how these could be used to support the findings of the study, the suggestions that were made in this study, the algorithm's formulation, and output, and verifying the interconnections between these. He also assessed the study's writing for consistency.

4.6.5. Transferability

Bell and Bryman (2007) state that qualitative findings primarily focus on the contextual distinctiveness and relevance of the part of the social environment being examined since qualitative research usually entails the intensive study of a small group of people having traits. Given that they promote the variety of perspectives offered by the participants on the subject under research, convenience and purposeful sampling were used as strategies to enhance the transferability of the study.

4.7. Chapter summary

The research methods and design that were employed in this study have been covered in this chapter. The theoretical and conceptual framework of EDM and LA served as the basis for this study and the foundation for the research design and methods. The qualitative research approach is determined to be the best suitable for this study based on its characteristics. Semi-structured interviews were used as qualitative research to conduct the study. The selection of participants and the methods and processes used for data analysis were outlined and justified in this presentation. Additionally, the ethical issues for this study were discussed in this chapter. The outcomes of this qualitative research are presented in depth in Chapter 5.

Chapter 5: Data presentation, analysis, and discussion of the findings

"What we find changes who we become." – Peter Morville.

5.1. Chapter overview

In this chapter, there is a critical discussion that focuses on the findings that were obtained from the participants in the study, intending to answer the research question: "*What tool do IT teachers require to be notified of programming concepts in which learners are having difficulty and to intervene?"* alongside its subsequent questions that are as follows:

1. How and when do teachers intervene with learners facing programming difficulties?
2. How effective is the method in preparing learners' understanding of the concept?
3. What effective tool informs IT teachers about issues in which learners are having difficulty?

The responses provided by the participants are used in this section to answer these sets of questions. To present these responses, seven categories generated two prevalent themes identified in the analysis of the interviews in Chapter Four, Table 4, which will be used to discuss this chapter. These categories are the utilisation of the CAPS document to structure learning objectives, integration of a planning process, the inclusion of real-life problems, collaborative learning, solution development, guidance by the teacher and measuring learner experiences. Each category is discussed in detail with relevant excerpts from the interview transcripts and literature. In addition, the categories are used to inform that there are two themes that the study generates to address the purposes of the study. The themes that were found are (1) The role of problem-based learning in programming, as it was found in the instructional method that IT teachers use in their programming classrooms; (2) Automated feedback tool, which can be depicted as a primary theme that runs through the study, where its main purpose is to inform Grade 10 IT teachers about concepts that they need to focus on, based on previous results using a supervised learning algorithm.

5.2 Generating categories

5.2.1. Utilisation of the CAPS document to structure objectives

For several years, learning objectives have been portrayed as the bedrock of every educational activity, and they have been exalted as the consequence of both the teaching and learning process. Learning objectives are clearly outlined so that the learner understands what to expect. All four participants stated that the CAPS document guides them in teaching and learning programming concepts. As mentioned in Chapter 1, programming falls under Solution Development in IT, where the creation of programming is a systematic, organised process that relies on logical thinking to solve computational problems, including data-related ones. Articulated below are two quotations from participants on how they use the CAPS document to structure their learning objectives:

> *I use the Annual Teaching Plan (ATP). In Grade 10, we work with basic Delphi. So, I need them to learn basic Delphi and be comfortable with both the interface and the developer. We only do basic amounts and basic structures. We do the IF, ELSE, CASE, and nested loops, and we will do the text files if we get some time.*

Teacher B echoes the same statement by guiding what learners should achieve by providing a practical strategy.

> *The learners should be able to produce a working programme. They should be able to use different concepts, from loop structures to databases to text files, and they need to integrate all of those working classes, objects, and classes into a working code. So, I follow the CAPS document, but because I teach IT as an extra subject, I only sometimes stick to the order in which it is prepared if I have covered all the work from grades 10 to 12.*

5.2.2. Integration of a planning process

Programming is fundamentally a process-oriented methodology. Problem-solving requires forethought, during which the learner chooses sufficient means related to the learning objective and implements an approach to learning success (Cho *et al.*, 2021). Through such a process, learners can think objectively and develop a systematic plan to tackle the problem and take the necessary steps. A fundamental aspect of achieving this in IT, ensuring a successful programming code, is embellishing the learners' algorithmic thinking skills. According to DoE (2011), this is accomplished by teaching learners how to build algorithms, solve problems, and write programmes using real-world examples. One of the participants applies this theory to his learners in detail so they can identify the types of challenges that call for specific algorithms.

> *To teach algorithmic thinking, I relate to everyday examples. Recently, I taught them the six steps in the programming process because they need to consider these steps when solving a problem. (1) Define the problem. (2) You need to understand the problem. (3) Decomposition, essential in computational thinking, means breaking down that problem. (4) You plan the logic of the programme. (5) You must use your solution development*

tools, IPO [Input Process and Output], trace tables, and flowcharts. (6) Produce an algorithm to solve a problem.

This quotation suggests that a learner can use related concepts to develop algorithms for various issues or circumstances by exploring algorithms to solve fundamental problems using these steps, which are essential in developing algorithmic thinking.

5.2.3. The inclusion of real-life problems

To effectively fix issues using programming code, learners must research and analyse problems in several settings (such as those related to science, technology, the environment, and daily life) (DoE, 2011). Programming that incorporates actual problems focuses on the attributes of the problem; the issues used in programming should be challenging while also being realistic and relevant to specific circumstances. Following the evidence, the participants used a comparable strategy for choosing a real-world issue that is important and relevant to learners' interests and provides helpful programming knowledge while simultaneously immersing learners in the learning process and helping them gain a greater understanding of programming.

Teaching programming is incredibly challenging. I am concerned that they do not have prior knowledge when they choose the subject in Grade 10. One of the ways to make it enjoyable is gaming. I sometimes get the feeling when the kids think that when they can be good gamers, they can be good programmers. So, I make it extremely exciting for them, especially working where they are starting. They work with the properties and components of Delphi. First, I ask them to insert a form, change the colour, height, and width, and make those visual changes. You can see, get them excited. From there onwards, we move into coding.

While Teacher D used gaming models to contextualise programming by allowing learners to use the Delphi IDE as a ground for the foundation for programming, Teachers A, B and C used real-life examples to enhance learners' understanding of problem-solving in programming:

Yesterday, I was teaching them about WHILE loops. So, I make relatable examples using a WHILE statement. Like WHILE you are making a cake requires many steps, you keep on adding the ingredients until it is suitable for baking. WHILE the oven is baking the cake, you wait for the time.

Like Teacher A, the notion of developing algorithmic thinking, Teacher C, necessitates the use of everyday examples for learners to adopt computational thinking.

I use everyday examples for them like before you cook, you know that you need to follow the recipe first. To do this, you need to follow this recipe up until the last step. So, then I ask them, how do you bake the cakes? What is it that you need? Then you ask them to mention all those steps, which will inform the algorithm course. So those are the example that I can use next.

When organising activities, Teacher B employs problem-solving activities in programming based on what is going on in the country's socio-economic environment.

Regarding problem-solving, I structure our questions for that lesson around whatever is topical. For example, in an election season, I make my grade 10s write a programme to do online voting and record the votes. So, for example, when we were having problems with the license system with people struggling to log online to book a driver's license test, we wrote a much smaller programme to do that.

From these exceptions, learning through problem-solving programming is driven by and structured around critical challenges that learners are likely to encounter in the real world. This results in learners developing programmes that solve those problems.

5.2.4. Collaborative learning

All participants recognised support and guidance in collaboration between learners, including the teacher, as a critical component of problem-solving in programming. Based on the evidence, the problem-solving programming approach that the participants used encouraged pairings or groups of learners to work together to integrate new knowledge into knowledge acquired. An instance of this is seen in the quotation below:

When teaching a new programming concept, one learner will immediately catch it and understand it. Once he is done working on the programme, he is often willing to assist other learners that are slower. Also, I put them on pair work when they have a challenging programme. However, I get them closer by walking around to see as they develop programmes. I can see how they are either succeeding or not.

Collaborative learning is a component that, according to evidence from Teacher A, significantly enhances learners' ability to solve programming problems and ensures their skill learning. Based on this quotation, collaborative learning is exhibited by the theory of social constructivism. According to Cho *et al.* (2021), in social constructivism, learners learn more in a group context, where they tend to exhibit greater levels of motivation and self-assurance.

5.2.5. Solution development

Creating a workable solution to the given challenge is known as solution development. The participants below suggest that solution development may be the most critical aspect of problem-solving in programming.

> *Learners should be able to produce a working programme. They should be able to use different concepts, from loop structures to databases to text files, and they need to integrate all of those working classes, objects, and classes into a working code.*

From the quotations, learners must exhibit modelled responses to the problem by developing a working programme, and they ought to get honest feedback for their effort. The learner can achieve higher degrees of competence in programming by translating foundational programming knowledge to demonstration (Cho *et al.*, 2021). In other words, learners should synthesise relevant information with past knowledge to produce a solution, and this activity is a valuable tool for retaining recently taught knowledge.

5.2.6. Guidance by the teacher

Another distinctive feature of programming classrooms relative to other instructional methods is the teacher's role as a facilitator. Assertions from the participants prompt that learners must cultivate self-direction to solve programming problems as the teacher's function shifts to that of a learning facilitator, mandating this improvement in the learners:

> *I do guide them, particularly those who are facing programming challenges. I sit with them and guide them. My role is to give them instructions.*

While Teacher B provides a practical example of being a facilitator in a programming class, Teacher C supplements this procedure with educational resources.

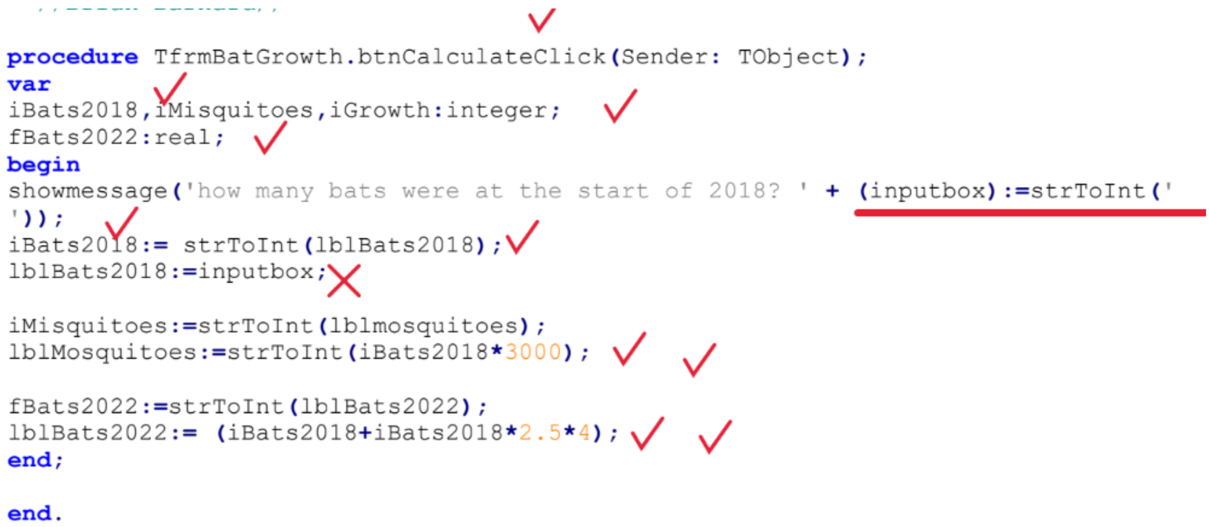
> *There is an eBook from MTN, which includes data files and guided activities. Let us say, for example, we go through a particular chapter, I make an example of the topic with them, and once we complete the example, I ask them to do a guided activity at home or in class.*

Evidence from the participants indicates that IT teachers have a crucial role in helping and evaluating the feasibility of learner solutions to a problem based on their expertise and knowledge of programming. The importance of the teachers' facilitation cannot be overstated. With it, learners can acquire vital assistance for developing a program at the right moment, developing correct codes, and showing enthusiasm for IT.

5.2.7 Measuring learning experiences

The evaluation of learning outcomes, among many other domains of education, is crucial in guaranteeing appropriate pedagogical quality and has consistently been challenging. According to Wang *et al*. (2012), the assessment of programming learning performance and experiences is complex since computer programming is problem-solving-oriented and extremely practical. Traditional evaluation methods are difficult to adapt to numerous innovations in computer programming education (Wang *et al*., 2012). Hence, the evidence below shows how IT teachers use contemporary teaching methodologies to improve learning results in programming.

> *I assess using homework, classwork, and formal assessments where they must develop their code. I use Adobe Reader to mark and highlight where they went correct or incorrect in their code.*



```
procedure TfrmBatGrowth.btnCalculateClick(Sender: TObject);
var
iBats2018,iMisquitoes,iGrowth:integer;
fBats2022:real;
begin
showmessage('how many bats were at the start of 2018? ' + (inputbox):=strToInt('
'));
iBats2018:= strToInt(lblBats2018);
lblBats2018:=inputbox;

iMisquitoes:=strToInt(lblmosquitoes);
lblMosquitoes:=strToInt(iBats2018*3000);

fBats2022:=strToInt(lblBats2022);
lblBats2022:= (iBats2018+iBats2018*2.5*4);
end;

end.
```

Figure 7: Marking guideline of a programming activity

Based on Figure 7, marking feedback is critical in the programming teaching and learning activities since it stimulates learner development and performance. Teacher B provides another measure of feedback to enhance the learners' progress.

> *In programming, I assess learners by requiring writing them to provide a solution for a programme. Sometimes they must build a programme from scratch, sometimes, it is a programme that is half-built, and they need to figure out what the solution is. Sometimes I will give them a set of notes to refer to, and then we will go through the programme in detail. The feedback is verbal feedback for informal tasks.*

Evidence from the quotation is that, like Teacher A, Teacher B requires learners to develop programmes and provides the use of verbal and notes feedback as a measure of learning experiences on how it improves learner performance by making them more aware of their progress toward achieving the anticipated proficiency in programming. The basis of this argument is like Teacher D.

> *So, in informal assessments, I use peer assessments and self-assessments. This teaches them how to go through each programme and work alone. Afterwards, I consult my memo with them and mark the task.*

While Teacher B and Teacher D shed light on how they use informal assessments to measure learner programming experiences, Teacher D provided enlightenment on how they measure learner experiences in the formal evaluation known as the Practical Assessment Task known as PAT. According to DoE (2011), the PAT is an application development assignment that allows learners to highlight their coding and application development abilities.

> *In formal assessments like Practical Assessment Task*[2]*, I use rubrics provided by the Department of Basic Education to score the learners' marks and provide feedback per the rubric.*

These rubrics have mark sheets that teachers can use to allocate their respective learners' marks based on the programmes they have developed (DoE, 2011).

Data analysis – Generating themes.

As mentioned in the preceding chapter, a theme is a repeating pattern or interpretation of the data that connects to the study topic. A theme is a more abstract concept requiring more data interpretation and integration than a category, which describes and organises the manifest content of a data collection (Kiger and Varpio, 2020). An inductive approach is used in this study to find the themes. These themes represent the inquiries made by participants and do not need to accurately reflect the researcher's interests or convictions (Braun and Clarke, 2012). As a result, an inductive approach is used to produce a deeper thorough study of the complete set of obtained evidence. The following thematic map enables the reader to analyse and contextualise findings based on two themes: The importance of problem-based learning in programming and there need for an automated feedback tool.

[2] A PAT score is a requirement of the final advancement mark for all learners studying subjects with practical components like IT. It contributes up to 25% of the final examination score.

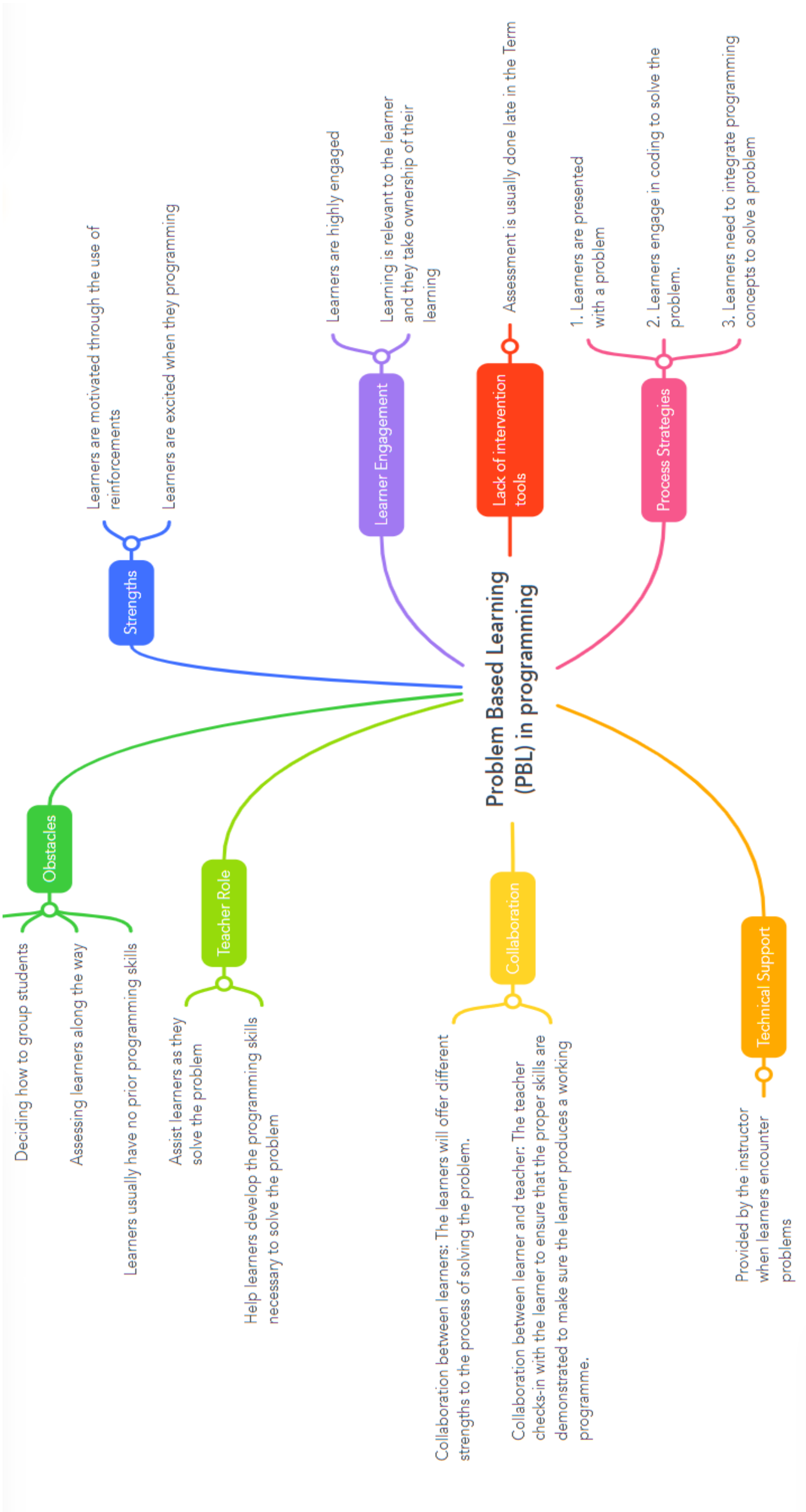


Figure 8: Thematic map

5.3. The role of problem-based learning in programming

Seven categories were examined in the previous section. Based on the substantive evidence provided by participants, the standard instructional method for teaching and learning programming concepts is problem-based learning (PBL). The evidence from participants shows that problem-based learning in programming is used to shape learning experiences. PBL is an educational technique that prioritises the needs of the learners and encourages active learning. PBL is based on the idea that the learner actively constructs knowledge throughout learning (Looi and Seyal, 2014). This means that learning is the consequence of a learner's actions; teachers only contribute by facilitating and supporting productive behaviours. Instead of teaching the learners about the problem and its solution, the teacher presents and helps them discover their solutions, guiding their inquiry and learning (Looi and Seyal, 2014). In a similar vein, Grade 10 IT teachers use programming with the aim of learners using their knowledge of the essential concept to solve substantive issues.

5.4. Automated feedback tool

Machine learning algorithms are one of the learner intervention tools that can be an advent. Teachers can take further steps to coordinate appropriate aid for the underperformers to advance in their subject and enhance their grades as a supplement to the feedback-giving methods (Lopez-Bernal *et al.*, 2021). The primary priority of the paper is to employ the supervised learning algorithm as an intervention tool to help IT educators identify programming concepts that learners are having trouble with early on and take the necessary steps to increase retention and performance. To accomplish this priority, previous IT learners' academic marks were applied to guide teachers about programming topics they should give heed to using a supervised learning algorithm. Using the frameworks provided by the EDM and LA, the analysis procedure used to create that output in the supervised learning algorithm is outlined below.

The process has three phases and similar procedures are conducted in each step (see Figure 9). The main effort is creating and preparing the training dataset (Swasti and Monika, 2013). Data cleansing and pre-processing fall under the category of data preparation. Cleaning data entails removing pointless features and dealing with cases when parts lack value. Data preparation enhances the dataset's quality even further, enabling the algorithms to provide better results (Swasti and Monika, 2013). Over the produced dataset, the supervised learning algorithm is run during the experimental assessment phase. A model that looks to be easier to comprehend is picked for interpretation. Therefore, the chosen model is put into a format that the teacher can understand based on the decision-tree algorithm.

Figure 9: Data steps

5.4.1 Data preparation

Data preparation is the first step, which involves visualising, cleaning, and unifying the data to prepare it for the next step (Khalaf *et al.*, 2016). Data preparation is crucial since data is erroneous, imperfect, and noisy (Swasti and Monika, 2013). Data may not be sufficient because it lacks desirable characteristics or may not be relevant to the study's topic. It is hence dirty as a result. Data preparation is crucial because dirty data leads to duplicate records, insufficient quality data, and—most crucially—poor results (Swasti and Monika, 2013). Quality data must serve as the foundation for quality judgments. Quality data must be consistently integrated into data warehouses. Data preparation allows for the measurement of data quality in terms of correctness, comprehensiveness, consistency, timeliness, plausibility, and interpretability through detailing three steps which include data description, data cleaning and data pre-processing (Swasti and Monika, 2013)

5.4.1.1 Data description

The dataset in this study consists of the learners' records for their scores on the Grade 10 IT final examination practical paper. The data was gathered from Grade 10 IT teachers, and all ethical requirements were strictly observed. The training dataset consists of programming topic areas and their concepts covered in Grade 10. The training dataset holds 12 cases, each with a maximum of four determination features and one estimation class (see Table 8).

Table 8: Topic areas of programming and their concepts

Programming topic area and their concepts	Programming concept ID
Procedural programming	
• Writing conditional statements • Solving mathematical problems in programming • Syntax • Understanding data types	• CS • SM • SYT • DT
Programme design	
• Composition of properties and components • Conceptualising problems and designing solutions • Debugging and exception handling	• COM • CDP • DE
Algorithm design	
• Abstraction/Pattern Recognition • Modelling of algorithms – Flowcharts and designing tables	• AP • MA
Object-oriented programming	
• Differentiating between classes and objects • Scope design of variables • Polymorphism	• OC • DV • PM

5.4.1.2 Data cleaning

Given that they are related to the learner's privacy, unnecessary features were not included in identifying the programming concepts that learners need help with. Removing excessive characteristics, such as the learner's name, gender, and the school's name, is done to abide by ethical and privacy considerations. Likewise, machine learning algorithms cannot effectively grasp and interpret noisy input (Khan *et al*., 2021). Unreliable or deceptive values for one or more features are considered noisy data. Such an occurrence degrades the algorithm's performance; hence these features were removed. Subsequently, the training dataset had 12 instances after eliminating the noisy data.

5.4.1.3 Data pre-processing

A substantial number of features make up the training dataset in most cases, yet, employing all of them may degrade the classification outcome (Khalaf *et al*., 2013). It is preferable to decide on a subset of relevant characteristics throughout the classification process. The pre-processing stage minimises complexity, improves computational effectiveness, and prevents overfitting by making the model comprehensible and less complicated (Khan *et al*., 2013). Several feature selection algorithms are available for this; however, in this study, the Gain Ratio Attribute Evaluator Filter with Ranker strategies was employed to reduce the number of overlapping features and rank programming concepts under the specific topic area from high to low. The learners' cumulative marks were normalised using a conversion process that changed the numerical values into nominal values. Table 9 displays a class label that divides programming ideas into high marks, medium marks, and low marks categories depending on the learner's grades.

Table 9: Class label

Class Label	**Description**	**Interval values**
L	Low marks	0-39
M	Medium marks	40-69
H	High marks	70-100

5.4.2 Experimental assessment

5.4.2.1 Algorithm development

As mentioned in Chapter 4, the WEKA tool is used to classify the programming concepts teachers must focus on. After pre-processing, the data was imported into the Weka tool. After the dataset has been pre-processed, the second step entails exporting the data (Khor, 2018) into the WEKA software to test it in a decision-tree algorithm. To choose the optimal method to divide the data, the decision-tree algorithm employs an internal measure to discover the split that leads to the most significant gain in forming accuracy (Mesarić and Šebalj, 2016). The training dataset was based on the topics of programming language covered in the IT subject and was classified against the testing dataset, which was the marks that learners obtained into a decision-tree algorithm.

5.4.2.2 Algorithm execution

Weka Explorer was opened with the ITmarks.artff file that had been produced. Based on the learners' performance, this study developed a model using four attributes to highlight programming ideas that teachers should focus on. These attributes fell under procedural programming, object-oriented programming, algorithm design, and programme design. The model's performance is evaluated using a confusion matrix after it has been trained and confirmed using a 10-fold cross-validation procedure. An exact solution to the classification issue is proven by a confusion matrix, which holds data on accurate classifications (Khor, 2018). Therefore, the J48 is used as the decision tree to stand for the algorithm to implement the supervised learning algorithm.

5.4.3 Algorithm's implementation

5.4.3.1 Algorithm's interpretation

Lastly, the output derived from the algorithm is used to analyse the results (Khalaf *et al.*, 2016). In this regard, a set of options is derived from the model to show educators the topics they will need to pay attention to and implement new teaching and learning strategies in their present learners (see Figure 10).

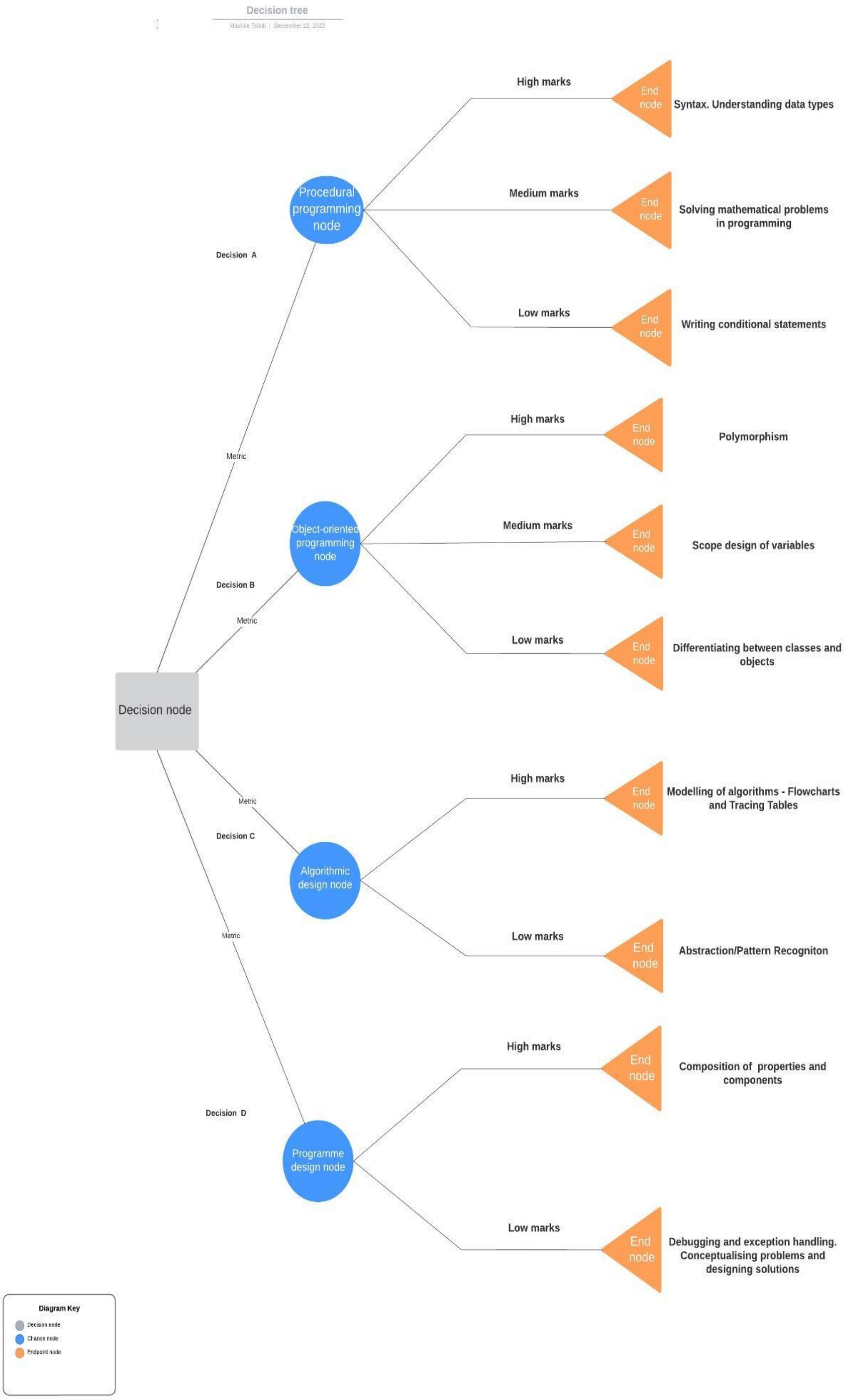


Figure 10: The algorithm's outcome

5.5 Discussion

Programming is used to teach logic and reasoning (Koorsse *et al*., 2015); hence it is efficient and well-suited to the PBL approach. Evidence from the participants demonstrates that through meticulously planned instruction, IT teachers aim to improve learners' ability to understand and deal with real-world problems through programming. One of the critical exit learning outcomes of CAPS, problem-solving abilities are a crucial qualification for citizenship in today's culture (DoE, 2011). Through resolving learning problems or real-world issues, individuals must acquire the fundamental idea and method of problem-solving, whether in life or at work. The need for problem-solving abilities in social issues is growing due to the move toward an information society (Koorsse *et al*., 2015). The contemporary information age can only be adopted through teaching methods that enable learners to collaborate to learn and be free to think and improve their interaction abilities. Unsurprisingly, the PBL approach perfectly meets this stipulation in IT subjects. Programming classrooms are learner-centred; bearing practical problems allows teachers to join their learners in groups and address issues by stimulating real-life situations through programming. Therefore, improving learners' problem-solving skills and making knowledge and meaning construction clear during the problem-solving process in programming.

It is challenging for newbie programmers like Grade 10 IT learners to comprehend these concepts without previous programming knowledge (Koorsse *et al*., 2015), so learning programming would become less appealing. Thus, IT teachers need to foster an environment that encourages learning and is entertaining for learners. Hence, their lessons are learner-centred, and IT teacher plan lessons to promote learners' abilities and alter the focus of every class to what should be learnt and how learning occurs in enhancing learners' interest in education and transform the programming learning process from passive to active engagement.

Teachers start by making a collection of engaging and intriguing programmes and demonstrating the dynamic effect on learners to pique learners’ interest immediately in programming classes. Learners’ curiosity is quickly piqued as it gives them the sense that they can put up a good multimedia show in just a few minutes (Peng, 2010). After that, the programming teaching and learning process is centred on challenges to pique learners' attention and ignite their enthusiasm for programming. Subsequently, it stimulates their thirst for more programming, inspires their creative side, and enables them to study with questions in mind from start to finish, solve problems, and apply their knowledge based on the programming concepts they have been taught.

As per the findings, learners' primary role in programming is emphasised as learner-centred, and a teacher's role shifts from only imparting knowledge to leading learners. This approach calls for teachers to have a firm grasp of the programming subject matter. Still, it also calls for flexible knowledge usage, extraordinary logical reasoning abilities, the ability to ask and answer questions, the capacity to pique learners' interest and enthusiasm for discussion, and the capacity to systematically arrange the learners while integrating activity and silence while solving problems (Koorsse *et al.*, 2015). Within the spectrum of participants' responses, the teachers have content knowledge of the subject and classroom management skills to facilitate the learning process. When it comes to problem-solving with the use of the PBL, the teachers place much emphasis on the creation of open, thought-provoking, and demanding skills of learners, and provide them with ample time and room to learn and explore and encourage them to produce something original and creative, to solve a problem in programming. Furthermore, the PBL approach helps learners develop their interpersonal and mutual aid skills through one-on-one, pair, and group interactions that allow them to engage in discussions to solve a problem.

To measure programme learning experiences, teachers use measure using informal and formal assessments. Evidence suggests and emphasises the importance of rigorous and process evaluation. According to Peng *et al*. (2010), it is a thorough assessment of learners' abilities and represents the calibre of every facet of a learner during the problem-solving process. An example of how the evaluation is completed is seen in Figure 7, where an assessment is based on the development of a learner's overall programming proficiency. This signifies that the learner's ability to solve issues by using what they have already learned in programming through individual study or collaborating with others to tackle practical difficulties is at the core of this evaluation. Hence, the teacher participants asserted that they provide constructive feedback to their learners. Constructive feedback is essential in teaching and learning because it serves as a stimulus for learners' performance and growth (Pangastuti *et al.*, 2022). Subsequently, feedback can improve learners' programming effectiveness by making them more conscious of their development toward achieving the desired competency.

Using the record of marks of IT learners, a supervised learning algorithm is used to identify programming concepts that IT learners based on data collected from four schools. The output of this information can assist teachers in finding initiate intervention programmes to fulfil the learners' learning needs based on the programming concepts they are struggling with, provide customised learner-centred experiences, and improve learning results (Nafea, 2018). Figure 10 shows the decision tree constructed from the ITmarks.artff file. The decision tree commences with the decision node, indicating a choice that needs to be taken—followed by four nodes. Each node represents one of four programming topic areas: procedural programming, programme design, algorithm design, and object-oriented programming. Afterwards, there are more branches, each of which denotes a potential high, medium, and lower mark based on each topic area of programming. In response to high, medium, and low marks, the endpoint node categorises programming concepts that teachers should note in each topic area.

These are the rules that were established for each topic area to categorise programming concepts:

1. IF ITMarks<=39 THEN Class = L
2. IF ITMarks>=40-69 THEN Class = M
3. IF ITMarks>=70-100 THEN Class = H

Table 10 shows the results obtained in the confusion matrix. Out of 12 instances, five are classified as 'L'. Which consists of five programming topics: Writing conditional statements, conceptualising problems, and designing solutions, Debugging and exception handling, Abstraction/Pattern Recognition, and differentiating between classes and objects, being displayed as programming concepts that learners are underperforming in. Hence, the true positive (TP) and false positive (FP) are 0.845 and 0.075, respectively. For the programming concepts that learners are performing averagely in, 2 out of 12 are classified as 'M': Solving mathematical problems in programming and Scope design of variables. Therefore, the TP rate is 0.732, and the FP rate is 0.25 (see Table 11). There are 5 out of 12 programming concepts that learners highly perform in. These are Syntax, understanding data types, Polymorphism, Modelling of algorithms – Flowcharts and designing tables, and the Composition of properties and components. They are classified as H with 0.61 TP and 0.86 FP rates (Refer to the programming concepts ID in Table 8).

Table 10: The confusion matrix of three labelled classes

	Labelled Class			
	Topic area	High marks	Medium Marks	Lower Marks
Actual Class	Procedural programming	DT, SYT	SM	CS
	Programme design	COM		CDP, DE
	Algorithm design	MA		AP
	Object-oriented programming	PM	DV	OC

Table 11 depicts the class-wise accuracy.

Table 11: Class-wise accuracy of the three labelled classes

Class	**True positive rate**	**False positive rate**
H	0.61	0.86
M	0.732	0.25
L	0.845	0.075

5.6 Chapter summary

This chapter reported the study’s findings. The study's significant conclusions identified PBL as one of the primary approaches used for programme teaching and learning in IT classrooms for learners in grade 10. The research identified six categories as crucial to the use of PBL in programming classes - Utilisation of the CAPS document to structure learning aims, integration of a planning process, the inclusion of real-life problems, collaborative learning, solution development and guidance by the teacher. Moreover, teachers employ informal and formal evaluations to monitor programme learning experiences. After completion, the learners receive helpful feedback on their performance in this evaluation based on their overall programming ability growth. Given Chapter 1, assessments are usually done late. When it comes to helping learners comprehend a concept, teachers lack the necessary tools. To assist schools in preventing learners from misunderstanding the programming concept, a tool that evaluates learner performance in programming concepts is required (Islam, Mouratidis and Mahmud, 2021). Using a visual tool in the form of a decision tree to identify key concepts, this strategy is used in this study to give teachers a report on programming principles they should pay attention to. The tree output was reported, enabling teachers to learn about the programming concepts with which their learners are having difficulty and to intervene earlier. The significance of the study's results is outlined in the next chapter, along with suggestions for future research.

Chapter 6: Conclusions

"The heart and soul of good writing is research; you should write not what you know but what you can find out about." – Robert J. Sawyer.

6.1. Chapter overview

Three specific purposes propelled this study. The first was to unbiasedly study and examine IT teachers for Grade 10 on how they shape educational experiences when they educate learners about programming principles. The second was the provision of literature on tried-and-true machine learning-based approaches that can inform teachers about the concepts or topics their learners are struggling with so that they may intervene and give aid. The third goal of the reach was to develop a model and put into practice a supervised learning algorithm that pinpoints the programming challenges that learners are having. The foundation informed this overarching purpose of the two earlier goals. Regarding the study's objectives, the central question for this study has been: "What tool do IT teachers require to be notified of programming concepts in which learners are having difficulty and to intervene?" The study concentrated on three sub-questions to address this subject.

1. How and when do teachers intervene with learners facing programming difficulties?
2. How effective is the method in preparing learners' understanding of the concept?
3. What effective tool informs IT teachers about issues in which learners are having difficulty?

As discussed in Chapter 1, teachers at schools lack the necessary resources when it comes to helping learners understand a topic. To assist educators in preventing learners from misunderstanding the topic, a programme that assesses learner performance in programming concepts is required. This research focused on developing a unique automated supervised learning approach that combines LA and learner performance management to pinpoint weak areas of programming for learners. In this study, the automated tool was used to identify learners who need immediate support and alert teachers to assist them in better comprehending the concepts with which they are struggling. Islam, Mouratidis, and Mahmud (2021) pointed out that this automated programme can close the gaps by fusing the most recent learner data to generate helpful predictive and proactive analyses once it is in place.

To structure the supervised learning algorithm, EDM and LA were proposed as the theoretical and conceptual frameworks of the study, respectively. EDM, as a theoretical framework in this study, is focused on designing, studying, and executing automated ways to find patterns in sets of data in the form of learner marks that would be challenging for teachers to analyse concepts that learners are struggling with because of the enormous amount of information within which they exist (Mohamad and Tasir, 2013). LA is a conceptual framework to capture, measure, and report data on learners on concepts they are struggling with and their settings to comprehend and optimise programme teaching and learning and the environments in which it occurs (Papamitsiou and Economides, 2014). To report on the findings using these frameworks, a thematic analysis was used as the study's analytical framework to find patterns within the participants' interview responses. From this frame of reference, this comprehensive approach informed the use of the WEKA tool in this study as it is an essential component within the machine learning algorithms for EDM and LA tasks and allowed the generation of the algorithm's output.

An overview of the study's findings is presented in this concluding chapter. This chapter starts by discussing the study's findings. The summary of the findings, significance and limitations of this study are then examined. Finally, the chapter ends with suggestions for future research, emphasising the significance of more significant research based on machine learning algorithms in South African education.

6.2 Summary of the findings

This research achieved its objectives and answered its research questions as follows: Regarding the research questions, the study found that IT teachers intervene with learners facing programming difficulties by employing both informal and formal assessments to evaluate learners' programming abilities and reflect on their problem-solving skills. The study also showed that the PBL approach is a method that IT teachers use to develop learners' comprehension of the programming. In response to the primary research question, the study recommended an automated feedback tool in the form of a supervised learning algorithm to highlight programming concepts that teachers should pay attention to. The study's findings offered proof in support of using the J48 decision tree algorithm to create a model that would highlight important t programming concepts for teachers. In conclusion, the study discovered that the research objectives and questions were fully achieved, and the proposed tool can assist IT teachers in identifying learners' programming difficulties and intervening promptly. A summary table is provided below to facilitate the reader's understanding of the extent to which each research objective and question was achieved.

Table 12: Summary of the extent to which research objectives and questions were achieved

Research objective/question	Extent of achievement
Objective 1: Review existing literature on machine learning algorithms and their applications in educational settings, particularly in identifying learner difficulties with programming concepts.	Fully achieved
Objective 2: Conduct semi-structured interviews with Grade 10 IT teachers to explore how they facilitate learning experiences for programming concepts.	Fully achieved
Objective 3: Collect and analyse data on learners' recent exam results to determine their performance in programming concepts.	Fully achieved
Objective 4: Develop and implement a supervised learning algorithm that identifies programming concepts that learners are struggling with.	Fully achieved
Question 1: What tool do IT teachers require to be notified of programming concepts in which learners are having difficulty and to intervene?	Fully achieved
Question 2: How and when do teachers intervene with learners facing programming difficulties?	Fully achieved
Question 3: How effective is the method in preparing learners' understanding of the concept?	Fully achieved
Question 4: What effective tool informs IT teachers about issues in which learners are having difficulty?	Fully achieved

6.3. Significance and implications of the study

The current educational system in South Africa puts much emphasis on giving learners distinct kinds of knowledge (van Zyl, 2016). A learner's ability to retain or recollect previously taught knowledge measures that learner's competence. The primary issue with this procedure is that it needs to assess the level of learner comprehension. The education industry has lately used machine learning and other innovations to tackle the problem. One of the applications of machine algorithms was used in this study. The supervised learning algorithm was applied to flag programming concepts that teachers need to pay attention to; based on the results of the learners' marks. Using the results obtained in the decision tree on programming concepts that teachers need to focus on improves efficiency. It encourages them to adapt instruction to the requirements of learners. Hence, this study endorses machine learning algorithms in educational fields because they can be employed to flag concepts that learners face based on data, which can be difficult to gauge from simple observations in a traditional classroom. This enhances teachers' pedagogical efficiency to enhance learner comprehension, which improves learner performance (Webb *et al*., 2020).

6.4. Limitations of the study

Due to the study's narrow emphasis on informing the IT teachers for Grade 10 on programming concepts that they need to focus on that, its ability to be generalised is constrained to the Grade 10 IT teachers (however, Grade 11-12 teachers and researchers can use this information on foundational programming concepts that IT learners are facing challenges with). This means that the findings reported in this research on programming concepts teachers need to pay attention to cannot be applied to another subject. This signifies that if the analysis is conducted on a similar topic, a distinct algorithm based on machine learning should be developed to prove learners' difficulties in a specific subject.

6.4. Future recommendations

Future research can focus on developing machine learning classification algorithms that can be applied in the classroom to forecast learners' final-year performance in Grade 12. Based on various academic factors, a supervised learning algorithm as a predictive technique can estimate the matric results for each IT learner. This study can be conducted quantitatively whilst considering the predictive nature of the algorithm, and positivism could be used as a research paradigm. Consequently, the algorithm's primary output lists the learners' likelihood of succeeding in the IT subject about their grades. Once the algorithm's work has been recognised, supplementary teaching and learning intervention tools may be provided to teachers and their respective learners, which may help learners achieve better results on the IT subject.

References

Aalst, W. V. (2016). *Data science in action. In Process mining.* Berlin, Heidelberg.: Springer.

Akhtar, D. M. (2016). Research design. Research Design (February 1, 2016). *Research in Social Science: Interdisciplinary Perspectives*, 68-84.

Alenezi, H. S., & Faisal, M. H. (2020). Utilizing crowdsourcing and machine learning in education: Literature review. *Education and Information Technologies, 25(4)*, 2971-2986.

Arifin, S. R. (2018). Ethical considerations in qualitative study. *International Journal of Care Scholars, 1(2)*, 30-33.

Babbie, E., & Mouton, J. (2007). *The practice of social research.* London: Thomson Wadsworth.

Baškarada, S., McKay, T., & McKenna, T. . (2013). Technology deployment process model. *Operations Management Research, 6(3)*, 105-118.

Bell, E. &. (2007). The ethics of management research: an exploratory content analysis. *British Journal of Management, 18(1)*, 63-77.

Bell, E. B. (2007). *Business research methods.* Oxford: Oxford university press.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19(1)*, 205-220.

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *Journal of the Learning Sciences, 22(4)*, 564-599.

Bhaskar, R., & Hartwig, M. . (2016). *Enlightened common sense: The philosophy of critical realism.* London: Routledge.

Bhutto, E. S., Siddiqui, I. F., Arain, Q. A., & Anwar, M. (2020). Predicting students' academic performance through supervised machine learning. *International Conference on Information Science and Communication Technology* (pp. 1-20). (ICISCT).

Blikstein, P., & Worsley, M. (2016). Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3(2)*, 220-238.

Bloomberg, L. D. (2019). *Completing Your Qualitative Dissertation (Fourth Edi).* California: SAGE Publications.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. E. Cooper, *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57-71). Washington: American Psychological Association.

Campbell, S. L., Jones, P. J., Williamson, G. J., Wheeler, A. J., Lucani, C., Bowman, D. M., & Johnston, F. H. (2020). Using digital technology to protect health in prolonged poor air quality episodes: A case study of the airrater app during the Australia. *Fire*, 3-40.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2(01)*, 20-28.

Cho, H., Cho, S. M., Cho, B. M., Jeong, S. H., & Cho, S. Y. (2021). A Study on Key Components of Problem-Based Learning through Literature Review. *Journal of Problem-Based Learning, 8(2)*, 69-74.

Coleman, C. B. (2019). A Better Cold-Start for Early Prediction of Student At-Risk Status in New School Districts. *International Educational Data Mining Society, 12(1)*, 732-737.

Cukurova, M. (2018). A syllogism for designing collaborative learning technologies in the age of AI and multimodal data. *European Conference on Technology Enhanced Learning* (pp. 291-296). Cham: Springer.

Education, D. O. (2008). *National Curriculum Statement. Grades 10-12 (General). Learning Programme Guidelines. Information Technology.* Pretoria: Department of Education.

Education, D. O. (2011). *Curriculum and assessment policy statement, FET phase 10–12, Information Technology. National curriculum statement.* Pretoria: Department of Basic Education.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE open, 4(1), 2158244014522633*, 1-10.

Fox, S. &. Do, T. (2013). Getting real about Big Data: applying critical realism to analyse Big Data hype. *International Journal of Managing Projects in Business, 6(4)*, 739-760.

Galpin, V. C., & Sanders, I. D. (2007). Perceptions of computer science at a South African university. *Computers & Education, 49(4)*, 1330-1356.

Gazette, S. A. (2013). *Protection of Personal Information (POPI) Act (No. 4 of 2013).* Pretoria: Government Gazette.

Hasni, T. F., & Lodhi, F. (2011). Teaching problem solving effectively. *ACM Inroads, 2(3)*, 58-62.

Havenga, M., & Mentz, E. (2009). The school subject information technology: a South African perspective. *In Proceedings of the 2009 Annual Conference of the Southern African Computer Lecturers' Association* (pp. 76 - 80). New York: Association for Computing Machinery.

Havenga, M., Mentz, E., Breed, B., Govender, D., Govender, I., Dignum, F., & Dignum, V. (2012). . (2012). A case study regarding teachers' problem-solving activities and approaches towards computer programming in diverse learning environments. *In 3rd International Conference on Society and Information Technologies ICTIT* (pp. 25-28). Orlando (USA): International Institute of Informatics and Systemics.

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review, 54(5)*, 3299-3348.

Islam, S., Mouratidis, H., & Mahmud, H. (2021). An automated tool to support an intelligence learner management system using learning analytics and machine learning. *In IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 494-504). Cham: Springer.

Ismail, M. N., Ngah, N. A., & Umar, I. N. . (2010). Instructional strategy in the teaching of computer programming: a need assessment analyses. *TOJET: The Turkish Online Journal of Educational Technology, 9(2),* 125-131.

Jacobs, C., & Sewry, D. A. (2010). Learner inclinations to study computer science or information systems at tertiary level. *South African Computer Journal,* 3-10.

Kabra, R. R., & Bichkar, R. S. (2017). Student's Performance Prediction Using Genetic Algorithm. *International Journal of Computer Engineering and Applications, 6(3),* 19-29.

Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers, 18(2),* 233-235.

Kazanidis, I. P. (2021). A learning analytics conceptual framework for augmented reality-supported educational case studies. *Multimodal Technologies and Interaction, 5(3), 9,* 1-16.

Kelleher, C., & Pausch, R. (2005). Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. *ACM Computing Surveys (CSUR)*, 83-137.

Khalaf, A., & Aizenstein, H. J. (2016). Studying depression using imaging and machine learning methods.*NeuroImage: Clinical, 10,*, 115-123.

Khan, I., Ahmad, A. R., Jabeur, N., & Mahdi, M. N. (2021). A Conceptual Framework to Aid Attribute Selection in Machine Learning Student Performance Prediction Models. *International Journal of Interactive Mobile Technologies, 15(15),* 4-19.

Khor, E. T. (2018). Mining educational data to predict learners' performance using decision tree algorithm. *Proceedings of the 9th Annual International Conference on Computer Science Education: Innovation & Technology* (pp. 101-104). Sydney: Global Science and Technology Forum Pte. Ltd.

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical teacher, 42(8),* 846-854.

Koorsse, M., Calitz, A. P., & Cilliers, C. C. (2010). Programming in South African Schools: The Inside Story. *Annual Conference of the South African Lecturer's Association (SACLA) 2010 Conference* (pp. 71-77). Warmbaths: University of Pretoria.

Koorsse, M., Cilliers, C., & Calitz, A. (2015). Programming assistance tools to support the learning of IT programming in South African secondary schools. *Computers & Education, 82,* 162-178.

Kovacic, Z. (2010). Early prediction of student success: Mining students' enrolment data. *Proceedings of Informing Science & IT Education Conference* (pp. 648-665). Wellington: Open Polytechnic.

Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research, 57(2)*, 448-470.

Loh, J. (2013). Inquiry into issues of trustworthiness and quality in narrative studies: A perspective. *Qualitative Report, 18(33)*, 1-15.

Looi, H. C., & Seyal, A. H. (2014). Problem-based learning: An analysis of its application to the teaching of programming. *International Proceedings of Economics Development and Research, 70*, 68-75.

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2021). Education 4.0: teaching the basics of KNN, LDA and simple perceptron algorithms for binary classification problems. *Future Internet, 13(8), 193,* 1-14.

Luan, H., & Tsai, C. C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society, 24(1),* 250-266.

Marshall, B., & Geier, S. (2020). Cross-disciplinary faculty development in data science principles for classroom integration. *SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 1207-1213). Portland: Association for Computer Memory.

McNiff, J. (2013). *Action research: Principles and practice.* London: Routledge.

Mentz, E., Bailey, R., Havenga, M., Breed, B., Govender, D., Govender, I., Dignum, F. and Dignum, V. (2012). The diverse educational needs and challenges of Information Technology teachers in two black rural schools. *Perspectives in education, 30(1),* 70-78.

Merceron, A., & Yacef, K. (2014). Mining student data captured from a web-based tutoring tool: Initial exploration and results. *Journal of Interactive Learning Research, 15(4)*, 319-346.

Merriam, S. B. (2009). *Qualitative Research: A Guide to Design and Implementation.* San Francisco: CA: John Wiley & Sons.

Mesarić, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review,* 367-388.

Mohamad, S. K., & Tasir, Z. (2013). Educational data mining: A review. *Procedia-Social and Behavioral Sciences, 97,* 320-324.

Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education & Computer Science, 8(11)*, 36-42.

Mukuna, K. R., & Aloka, P. J. (2020). Exploring educators' challenges of online learning in COVID-19 at a rural school, South Africa. *International Journal of Learning, Teaching and Educational Research, 19(10)*, 134-149.

Nafea, I. T. (2018). Machine learning in educational technology. In H. Farhadi, *Machine learning-advanced techniques and emerging applications,* (pp. 175-183). London, United Kingdom: IntechOpen.

Nasteski, V. (2018). *USING MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION TO IMPROVE PERFORMANCE OF BIG DATA PROCESSING (КОРИСТЕЊЕ НА АЛГОРИТМИ ЗА КЛАСИФИКАЦИЈА НА ГОЛЕМИ ПОДАТОЦИ ЗА ПОДОБРУВАЊЕ НА ПЕРФОРМАНСИТЕ ПРИ НИВНА ОБРАБОТКА).* Bitola: St. Kliment Ohridski University.

Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business, 10(1)*, 3-12.

Pangastuti, D., Widiasih, N., & Soemantri, D. (2022). Piloting a constructive feedback model for problem-based learning in medical education. *Korean Journal of Medical Education, 34(2)*, 131-143.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society, 17(4)*, 49-64.

Pathan, S. A., Bhutta, Z. A., Moinudheen, J., Jenkins, D., Silva, A. D., Sharma, Y., ... & Thomas, S. H. (2016). Marginal analysis in assessing factors contributing time to physician in the Emergency Department using operations data. . *Qatar Medical Journal*, 1-17.

Pears, A., Seidman, S., Malmi, L., Mannila, L., Adams, E., Bennedsen, J., & Paterson, J. (2007). A survey of literature on the teaching of introductory programming. *ITiCSE-WGR '07: Working group reports on ITiCSE on Innovation and technology in computer science education* (pp. 204-223). New York: Association for Computing Machinery.

Peng, J., Yuan, B., Spector, J. M., & Wang, M. (2019). Integrating technology in programming learning and instruction: a critical review. *International Journal of Smart Technology and Learning, 1(4)*, 323-343.

Peng, W. (2010). Practice and experience in the application of problem-based learning in computer programming course. *International Conference on Educational and Information Technology (Vol. 1)* (pp. 1-170). Chongqing: IEEE.

Potter, G., Lopez.,. (2001). *In After Postmodernism: An Introduction to Critical Realism.* London and New York: Continuum.

Ramaswami, M., & Bhaskaran, R. (2009). A study on feature selection techniques in educational data mining. *Journal of Computing, Volume 1, Issue 1,* 7-11.

Rashid, Y., Rashid, A., Warraich, M. A., Sabir, S. S., & Waseem, A. (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods, 18,* 1-13.

Rist, R. S. (1996). Teaching Eiffel as a first language. *Journal of object-oriented programming, 9(1),* 30-41.

Romero, C., Venture, S., Garcia, E. (2007). Data mining in course management systems: Moodle case study tutorial. *Computers & Education, 51,* 368-384.

Salihoun, M. (2020). State of art of data mining and learning analytics tools in higher education. *International Journal of Emerging Technologies in Learning, 15(21)*, 58-76.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2(3),* 1-21.

Sayer, A. (2004). Why critical realism. In S. &. Fleetwood, *Critical realist applications in organisation and management studies (Vol. 11)* (pp. 6-20). London: Psychology Press.

Scott, D., & Morrison, M. . (2006). *Key ideas in educational research.* London: Continuum.

Seymour, L., Hart, M., Haralambous, P., Natha, T., & Weng, C. W. (2005). Inclination of scholars to major in information systems or computer science: reviewed article. *South African Computer Journal, 2005(35),* 97-106.

Singhal, S. &. (2013). A study on WEKA tool for data preprocessing, classification and clustering. *International Journal of Innovative Technology and Exploring Engineering, 2(6),* 250-253.

Steiner, C. M., Kickmeier-Rust, M. D., & Albert, D. (2014). Learning analytics and educational data mining: An overview of recent techniques. *Learning Analytics for and in Serious Games, (6)*, 61-75.

Tan, M., Dos Santos, C., Xiang, B., & Zhou, B. (2016). Improved representation learning for question answer matching. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 464-473). New York: IBM Watson Core Technologies.

Thiede, K.W., Brendefur, J.L., Osguthorpe, R.D., Carney, M.B., Bremner, A., Strother, S., Oswalt, S., Snow, J.L., Sutton, J. and Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education, 49,* 36-44.

Tijani, F., Callaghan, R., & de Villers, R. (2020). An investigation into pre-service teachers' experiences while transitioning from Scratch programming to procedural programming. *African Journal of Research in Mathematics, Science and Technology Education, 24 (2),* 266-278.

Tolsgaard, M. G.-S. (2020). The role of data science and machine learning in Health Professions Education: practical applications, theoretical contributions, and epistemic beliefs. *Advances in Health Sciences Education*, 1057-1086.

Van Zyl, S., Mentz, E., & Havenga, M. (2016). Lessons learned from teaching Scratch as an introduction to object-oriented programming in Delphi. *African Journal of Research in Mathematics, Science and Technology Education, 20(2)*, 131-141.

Vandamme, J. P. (2007). Predicting academic performance by data mining methods. *Education Economics,* 405-419.

Varpio, L., Paradis, E., Uijtdehaage, S., & Young, M. (2020). The distinctions between theory, theoretical framework, and conceptual framework. *Academic Medicine, 95(7), 989-994.*, 989-994.

Waite, J., & Sentance, S. (2021). *Teaching programming in schools: A review of approaches and strategies.* Cambridge, United Kingdom: Raspberry Pi Foundation.

Wang, Y., Li, H., Feng, Y., Jiang, Y., & Liu, Y. (2012). Assessment of programming language learning based on peer code review model: Implementation and experience report. *Computers & Education*, *59(2)*, 412-422.

Webb, M. E.-S. (2021). Machine learning for human learners: opportunities, issues, tensions and threats. *Educational Technology Research and Development, 69(4)*, 2109-2130.

Weisfeld, M. (2009). *The object-oriented thought process.* New Jersey: Addison-Wesley: Upper Saddle River.

Wilson, C., Sudol, L. A., Stephenson, C., & Stehlik, M. . (2010). *Running on empty: The failure to teach k--12 computer science in the digital age.* New York: Association for Computing Machinery.

Appendices

Appendix A: Ethics clearance

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | WSoE Wits School of Education

SCHOOL OF EDUCATION ETHICS COMMITTEE

CONSTITUTED UNDER THE UNIVERSITY HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER:** 2022ECE038M
PROJECT TITLE	Using past Information Technology results to inform educators about the concepts that they need to pay attention to using a supervised learning algorithm.
INVESTIGATOR	Mashite Tshidi
SCHOOL/DEPARTMENT OF INVESTIGATOR	**WSOE**
DATE CONSIDERED	22 July 2022
DECISION OF THE COMMITTEE	Approved unconditionally
RISK LEVEL	No Risk
EXPIRY DATE	Date of submission of the Research Report
ISSUE DATE OF CERTIFICATE	**CHAIRPERSON** ____________ Dr. Batseba Mofolo-Mbokane

cc: Dr Alton Dewa

DECLARATION OF INVESTIGATOR

To be completed in duplicate and **ONE COPY** returned to the Chairperson of the School/Department ethics committee.

I fully understand the conditions under which I am are authorized to carry out the abovementioned research and I guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

____
Signature

09 / 09 / 2022
Date

27 St Andrews Road, Parktown, Johannesburg, 2193 | Pvt Bag 3, WITS, 2050 | **T** +27 11 717 3007 | www.wits.ac.za/education

IYUNIVESITHI YASEWITWATERSRAND | YUNIVESITHI YA WITWATERSRAND

Appendix B: Gauteng Department of Education research approval letter

GAUTENG PROVINCE
Department: Education
REPUBLIC OF SOUTH AFRICA

8/4/4/1/2

GDE RESEARCH APPROVAL LETTER

Date:	04 July 2022
Validity of Research Approval:	08 February 2022– 30 September 2022 2022/276
Name of Researcher:	Tshidi M
Address of Researcher:	616 Capricon Street
	Mashimong
	Tembisa
Telephone Number:	061 339 3346
Email address:	*1672247@students.wits.ac.za*
Research Topic:	Using past Information Technology results to inform educators about the concepts that they need to pay attention to using a supervised learning algorithm.
Type of qualification	Master of Education
Number and type of schools:	5 Secondary Schools
District/s/HO	Ekurhuleni South

Re: Approval in Respect of Request to Conduct Research

This letter serves to indicate that approval is hereby granted to the above-mentioned researcher to proceed with research in respect of the study indicated above. The onus rests with the researcher to negotiate appropriate and relevant time schedules with the school/s and/or offices involved to conduct the research. A separate copy of this letter must be presented to both the School (both Principal and SGB) and the District/Head Office Senior Manager confirming that permission has been granted for the research to be conducted.

04/07/2022

The following conditions apply to GDE research. The researcher may proceed with the above study subject to the conditions listed below are met. Approval may be withdrawn should any of the conditions listed below be flouted:

1

Making education a societal priority

Office of the Director: Education Research and Knowledge Management
7th Floor, 17 Simmonds Street, Johannesburg, 2001
Tel: (011) 355 0488
Email: Faith.Tshabalala@gauteng.gov.za
Website: www.education.gpg.gov.za

1. *The letter would indicate that the said researcher/s has/have been granted permission from the Gauteng Department of Education to conduct the research study.*
2. *The District/Head Office Senior Manager/s must be approached separately, and in writing, for permission to involve District/Head Office Officials in the project.*
3. ***Because of the relaxation of COVID 19 regulations researchers can collect data online, telephonically, physically access schools, or may make arrangements for Zoom with the school Principal. Requests for such arrangements should be submitted to the GDE Education Research and Knowledge Management directorate.***
4. ***The Researchers are advised to wear a mask at all times, Social distance at all times, Provide a vaccination certificate or negative COVID-19 test, not older than 72 hours, and Sanitise frequently.***
5. *A copy of this letter must be forwarded to the school principal and the chairperson of the School Governing Body (SGB) that would indicate that the researcher/s has been granted permission from the Gauteng Department of Education to conduct the research study.*
6. *A letter/document that outlines the purpose of the research and the anticipated outcomes of such research must be made available to the principals, SGBs, and District/Head Office Senior Managers of the schools and districts/offices concerned, respectively.*
7. *The Researcher will make every effort to obtain the goodwill and cooperation of all the GDE officials, principals, and chairpersons of the SGBs, teachers, and learners involved. Persons who offer their cooperation will not receive additional remuneration from the Department while those that opt not to participate will not be penalised in any way.*
8. *Research may only be conducted after school hours so that the normal school program is not interrupted. The Principal (if at a school) and/or Director (if at a district/head office) must be consulted about an appropriate time when the researcher/s may carry out their research at the sites that they manage.*
9. *Research may only commence from the second week of February and must be concluded before the beginning of the last quarter of the academic year. If incomplete, an amended Research Approval letter may be requested to conduct research in the following year.*
10. *Items 6 and 7 will not apply to any research effort being undertaken on behalf of the GDE. Such research will have been commissioned and be paid for by the Gauteng Department of Education.*
11. *It is the researcher's responsibility to obtain written parental consent of all learners that are expected to participate in the study.*
12. *The researcher is responsible for supplying and utilising his/her research resources, such as stationery, photocopies, transport, faxes, and telephones, and should not depend on the goodwill of the institutions and/or the offices visited for supplying such resources.*
13. *The names of the GDE officials, schools, principals, parents, teachers, and learners that participate in the study may not appear in the research report without the written consent of each of these individuals and/or organisations.*
14. *On completion of the study, the researcher/s must supply the Director: Knowledge Management & Research with one Hard Cover bound and an electronic copy of the research.*
15. *The researcher may be expected to provide short presentations on the purpose, findings, and recommendations of his/her research to both GDE officials and the schools concerned.*
16. *Should the researcher have been involved with research at a school and/or a district/head office level, the Director concerned must also be supplied with a summary of the purpose, findings, and recommendations of the research study.*

The Gauteng Department of Education wishes you well in this important undertaking and looks forward to examining the findings of your research study.

Kind regards

....

Mr. Gumani Mukatuni
Acting CES: Education Research and Knowledge Management

DATE: 04/07/2022

Office of the Director: Education Research and Knowledge Management
7th Floor, 17 Simmonds Street, Johannesburg, 2001
Tel: (011) 355 0488
Email: Faith.Tshabalala@gauteng.gov.za
Website: www.education.gpg.gov.za

Appendix C: Research proposal outcome

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

FACULTY OF HUMANITIES
POSTGRADUATE OFFICE

08 August 2022
Student Number: 1672247

Mashite Tshidi
616 Kaporikone Street
Mashemong Section Tembisa
Johannesburg 1632
Gauteng South Africa
By Email:
1672247@students.wits.ac.za
Cc: Supervisor

Dear Mr. Tshidi

APPROVAL OF PROPOSAL FOR THE DEGREE OF MASTER OF EDUCATION

I am pleased to be able to advise you that the readers of the Graduate Studies Committee have approved your proposal entitled "***Supervised learning algorithm's role in flagging Information Technology concepts that call for teachers' attention"***. I confirm that **Dr Alton Dewa** has been appointed as your supervisor in the School of Education.

The research report is normally submitted to the Faculty Office by 15 February, if you have started the beginning of the year, and for mid-year the deadline is 31 July. All students are required to RE-REGISTER at the beginning of each year

Please note that should you miss the deadline of 15 February or 31 July you will be required to submit an application for extension of time and register for the research report extension. Any candidate who misses the deadline of 15 February will be charged fees for the research report.

Please keep us informed of any changes of address during the year.

Note: All MA and PhD candidates who intend graduating shortly must meet your ETD requirements at least 6 weeks after your supervisor has received the examiners reports. **A student must remain registered at the Faculty Office until graduation.**

Yours Sincerely
Faith Herbert
Faith Herbert
Senior Faculty Officer
Faculty of Humanities

Private Bag 3, WITS 2050, South Africa I T + 27 11 717 3018 I E help.humanities@wits.ac.za I www.wits.ac.za/humanities
IYUNIVESITHI YASEWITWATERSRAND | YUNIVESITHI YA WITWATERSRAND

Appendix D: Participation letter

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

PARTICIPANTS' INFORMATION SHEET

30 June 2022

Dear Sir/Madam

My name is Mashite Tshidi; I am a Master of Education student at the University of the Witwatersrand, Johannesburg. Under the supervision of Dr Alton Dewa, I am conducting a research study titled: **Supervised learning algorithm's role in flagging programming concepts that call for Information Technology teachers' attention.**

I am inviting you to take part in an interview for my proposed research. If you decide to take part, your participation in this research study will last about 30-45 minutes. The interview/research activity will take place at your place of work after school. With your permission, I would like to audio record the interview and obtain documents that have previous examination marks. This data will be stored in a password-protected laptop for 3-5 years and deleted thereafter. Only the researcher will have access to the data.

The conversation will be kept private and anonymous. When I disclose the research study's findings, I will not include your name or any other information that could identify you. Other researchers may utilise the data generated from this research project with your agreement, but your identity and any personal information will not be used or shared.

Participation is done in volunteering terms. You are not required to participate. You are free to leave the study at any moment. If you don't want to, you don't have to answer any questions. If you agree to participate in the research study, you will not receive any direct advantages. If you choose not to join, you will not lose any services, advantages, or rights that you would otherwise have. It will not cost you anything to participate in the research study.

This research study will be written up as a research report. The report will be available on the university library website. If you would like to receive a summary of this report, I will be happy to send it to you.

If you have any questions regarding this research project while or after it is completed, please contact me or my supervisor using the information provided below. If you have any problems or complaints regarding the research study's ethical processes, please contact us. You are welcome to contact the University Human Research Ethics Committee (Non-Medical), telephone +27(0) 11 717 1408, email, hrecnon-medical@wits.ac.za.

Yours sincerely,
Mashite Tshidi

Researcher:
Mashite Tshidi
1672247@students.wits.ac.za
0613393346

Supervisor:
Dr Alton Dewa
alton.dewa@wits.ac.za
011 717 3337 -0726828991

Appendix E: Interview questions

Research Topic: Supervised learning algorithm's role in flagging programming concepts that call for Information Technology teachers' attention.

Interview Questions

1. What learning objectives do you seek to attain particularly concerning programming in IT?
2. How do you formulate learning experiences to attain those objectives?
3. How do you measure understanding concerning programming concepts?
4. What feedback do you use to support learners' understanding of programming concepts?
5. What tool do you use to check the learner's marks after an assessment?
6. Do you schedule time for reflection?
7. How do you see if learners are lacking knowledge or understanding of a concept and how do you go about it in resolving that?