MACHINE LEARNING MODEL FOR EDUCATION LEVELLING IN MULTICULTURAL COUNTRIES USING UAE AS A CASE STUDY

By

SHATHA RASHED GHAREEB

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

September 2022

DECLARATION

I, Shatha Rashed Ghareeb, declare that this thesis, submitted to Liverpool John Moores University to fulfil the Doctor of Philosophy requirements, has not been submitted to other universities and institutes. I confirm that the work described in this PhD thesis is my own except for some sources that support my research, which are appropriately cited and indicated.

Shatha Rashed Ghareeb

September 2022

ACKNOWLEDGEMENT

In the beginning, all glory is owing to the almighty ALLAH (SWT), the most gracious compassionate for granting me the determination and capability to complete my long journey with this PhD study.

I place on record and warmly acknowledge the timely suggestions, continuous encouragement, inspired guidance, and invaluable supervision of my supervisor Prof. Abir Hussain for her support and advice that helped me face any challenges successfully. I am so grateful for reviewing my writing skills, correcting the technical parts and valuable feedback on the thesis, which has led to its publication in notable journals and conferences. I want to express my gratitude to Prof. Dhiya Al-Jumeily and Dr Wasiq Khan for their constant help and support of my research with practical advice. Special thanks go to Dr Thar Baker for his advice, feedback, support throughout the first year and valuable input in constructing the Web-based system. I also owe an outstanding debt to Dr Mohammed Khalaf for his continuous support.

Many thanks go to all staff of the Computer Science Department at Liverpool John Moores University for their support and technical assistance with my work, especially Tricia Waterson.

Words cannot express my feeling toward my family and friends, who deserve great thanks for their encouragement and support during my study journey.

ABSTRACT

In multicultural countries, there are several education curricula and various age groups for starting education in each curriculum to meet the population's needs. We are discussing the Abu Dhabi case as an example of such an environment with more than 170 nationalities. Some students are currently not being placed in the correct year group when changing curricula due to distinct academic start and end dates, and age standards. As a result, gaps exist in students' learning and performance. In addition, students' data are not centralized and utilized throughout their academic journey. This research proposed and developed a computational framework for such scenarios using Machine Learning (ML) techniques to help predict the most suitable levels for students when transferring between curricula, assigning these levels automatically, and holding students' data throughout their academic journey. Students' datasets were collected from their educational records for two consecutive academic years to fulfil this goal, and then pre-processing techniques were applied to the raw dataset. The research focused on how machine learning can predict students' levels using several models including Artificial Neural Network and Random Forest, alongside assembled classifiers. Extensive simulation results indicated that the Levenberg-Marquardt Neural Network method (LEVNN) has the best average results among the other applied methods. A user-friendly platform has been designed based on a web-based student management system to bring both perspectives together in one platform for schools and parents. The research would help education providers predict students' correct levels more efficiently without regular examinations, saving time and cost for schools, students, and parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THESIS ACRONYMS

Abbreviation	Definition
AI	Artificial Intelligence
ML	Machine Learning
ANN	Artificial Neural Network
NN	Neural Network
BPXNC	Back-Propagation Neural Network Classifier
LEVNN	Levenberg-Marquardt Neural Network
LNN	Linear Neural Network
RFC	Random Forest Classifier
F1	F1 Score
J Score	Youden's J statistic (J Score)
ROC	Receiver Operator Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
KNN	k-nearest neighbours algorithm
SVM	Support Vector Machine
TPR	True Positive Rate
TFR	True False Rate
ADEK	Department of Education and Knowledge
MOE	Ministry of Education
IGCSE	British curriculum International General Certificate of Secondary Education
DR	Dimensionality Reduction
RS	Feature Selection

FE	Feature Extraction
AUC	Area Under the Curve
PPV	Positive Predictive Value
t-SNE	T-distributed Stochastic Neighbourhood Embedding
SPE	Stochastic Proximity Embedding
Cfs	Correlation-based feature selector
IG	Information Gain
SU	Symmetric Uncertainty

Chapter 1 Introduction

1.1 Background

Pupil's levelling is a process based on age that lacks automation and a unified approach amongst schools in the UAE. In this case, based on the schools' curriculum, they could offer pass/fail or credit/no credit or letter grades. Many countries provide a variety of school curricula to address the issue, including British, American, and Arabic, yet this can create a conflict when it comes to pupils' levelling after transferring curricula. Machine learning (ML) can offer an accurate levelling system for schools to determine student levelling [1]. Education institutes such as schools and colleges use ML to identify struggling students earlier and develop action plans to improve success and retention. ML is expanding the availability and effectiveness of online learning by having differentiated learning for each student's needs [2].

In multicultural countries with many different nationalities, various education curricula satisfy parents' and students' needs. Switching from one curriculum to another has become a critical issue as differences among curricula could create gaps in education levels for students. Hence, it is challenging to assign the right level for students when moving to a new curriculum. The aim is to make this transition more precise and smoother with minimum effect on students' performance. Each education curriculum in Abu Dhabi has different age groups when children start school. For example, some education curricula (e.g., Philippines) start from 2.5 years old, whilst others start from 4 years old (e.g., India and Pakistan). According to analyses of education curricula outcomes in Abu Dhabi and based on students' results at the high school level, most students, who have transferred from one education curriculum to another, have an education gap in their education skills and knowledge. There is a need to have a united computerised system for automating the levelling process for students when changing the most common education curricula. Nevertheless, this area has a vast ambiguity that creates conflicts between parents, schools, and the Ministry of Education (MOE). This conflict harms the students' performance when making the wrong decision.

This research will focus on designing and developing a framework that utilises the latest machine learning algorithms and distributed systems paradigms (cloud and fog computing) to facilitate the students' levelling across educational curricula. It can help decision-makers specify the correct level/stage when enrolling new students. With extensive research on machine learning in education and the conducted literature review, such a tool is currently unavailable within the UAE's educational context; therefore, it deems novel.

1.2 Research Problem and Challenges

In multicultural countries where many expatriates are found, several education systems accommodate parents' and pupils' needs. Switching from one system to another has become a critical issue as differences could create gaps in education levels for students. Hence, it is complicated to assign the right level for students when moving to a new curriculum. The challenge is to provide a smooth transmission with minimum effects on pupils' performance.

Some researchers discussed the complexities of moving between different education systems. They looked at the potential pitfalls and the timelines to meet to ensure a smooth transition; "Transitioning back into the UK education system is a very difficult proposition for many students," says education consultant Elizabeth Sawyer, of Bennett School Placement [3].

The age groups for each year vary amongst curricula. There is an instant need for having a united computerised system for automating the levelling process for students when changing between education systems. Nevertheless, there is considerable ambiguity in this area, which usually creates conflicts between parents, schools, and the Ministry of Education, which generally happens because there's no common levelling guide among schools.

1.3 Research Questions

The following research questions below are addressed in our thesis.

1- Which machine learning algorithm performs best in classifying student levelling datasets?
2- Are the classification performance evaluation metrics effective in handling missing values, data cleansing and normalising the unbalanced dataset?
3- How to analyse the collected raw data to develop a dataset?
4- How to test the usability and effectiveness of student data in machine learning models?
5- Can past student data be used to provide accurate student levelling predictions?

The research aims and objectives address these questions to discuss the above research questions.

1.4 Aims and Objectives

This research aims to propose and develop a computational framework applicable in multicultural countries, such as UAE, where multi-education curricula (i.e., UK and USA curricula) are implemented. The goal is to aid in a smooth transition during admissions, levelling and differentiation of students who relocate from one education curriculum to another and minimise switching on the student's educational performance. Several objects must be considered as stated below to achieve these aims:

1. To review works of literature on several education curricula applied in multicultural countries.
2. To review works of literature on the applicability of Cloud Computing, Fog Computing and Machine Learning.
3. To collect primary data related to the research problem from schools with different curricula.
4. To conduct exploratory data analysis and select the relevant features that would assist in levelling the students across different education curricula.
5. To develop ML and Rule-based framework that analyses the student's current/historical educational data to infer the best/new level.
6. To deploy ML algorithms and provide predictions on the most suitable level for the student.
7. To develop a rule-based system that integrates the rules of admission and the ML predictions to automate proper decisions of student levelling.
8. To test the proposed tool and viability of using these technologies via several experiments including stakeholders' opinions on the mechanism.

1.5 Research Framework

At the beginning of the research process, I will review the literature on student levelling when transferring between different curricula in multicultural countries, ML algorithms and techniques, and cloud and fog technologies. After gaining enough knowledge and the required information, I will start the process of getting the required approvals including the ethical approval and a letter of permission from the concerned parties. Then I will start collecting data from schools with different curricula, pre-processing this data using the optimal techniques and applying a feature selection method to get the best-related features for my research. An adequate analysis will be applied to promote the understanding of the collected data and get the optimum results. Then an ML-based and Web-based framework will be developed to provide a clear idea of how data will be prepared and processed to get the best prediction results for levelling students when transferring between different curricula. After that, I will deploy the ML system to the dataset and get the optimal prediction. In addition, I will develop a rule-based system to provide a smooth transition between schools. Integration of the chosen ML algorithm and the rule-based system will be developed to use the stored data in the cloud and get the ideal prediction for the candidate. Finally, I will be testing the system to evaluate and validate it and make sure that it adds value and provide solutions to the research problem. Figure 1.1 illustrates the research framework workflow.

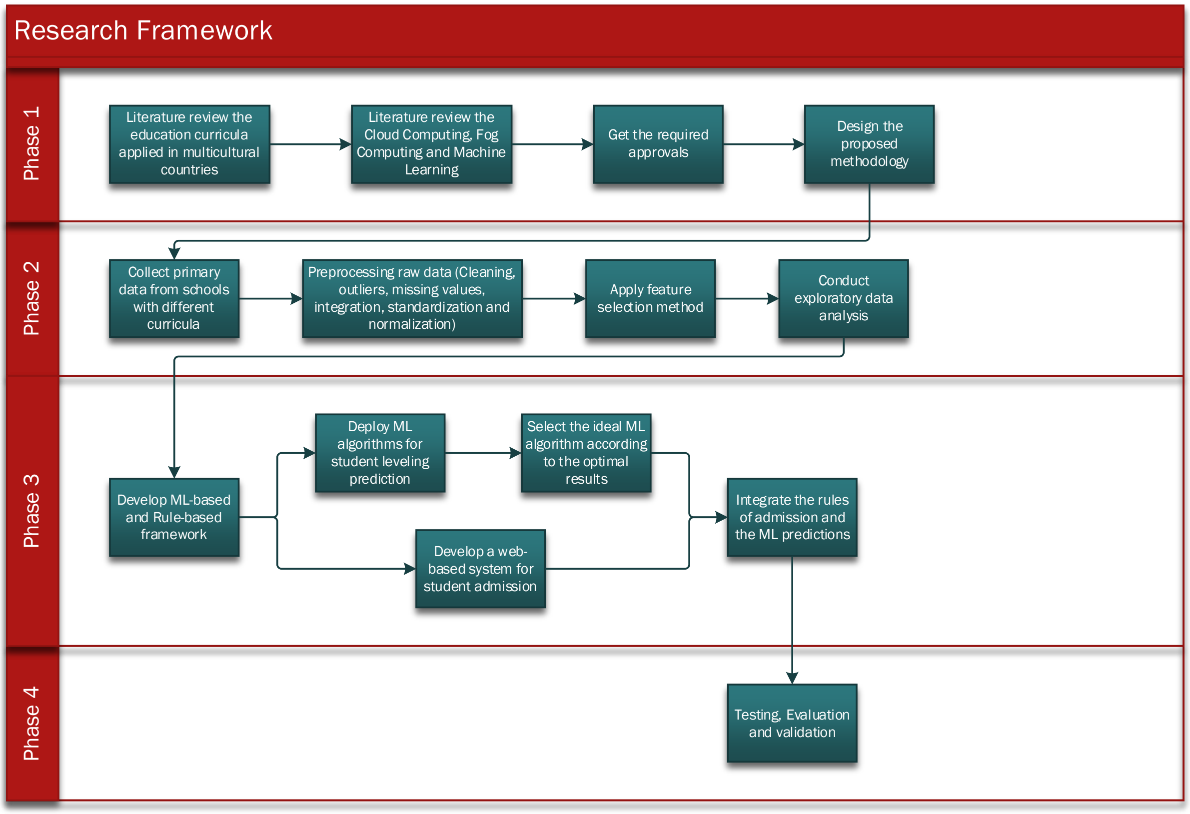


Figure 1.1 Research Framework

1.6 Research Contributions

The primary objectives of this research are to review the literature on various education curricula which are applied in multicultural countries. Study and critically discuss the challenges associated with the state-of-the-art levelling methods used at schools and governmental bodies (e.g., Ministry of Education) to understand the problem and determine the gap and shortcomings. Collect relevant data from education resources and stakeholders, and construct a baseline dataset by employing pre-processing, balancing, and filtering techniques. The constructed baseline dataset must contain sufficient samples for appropriate data modelling to conduct research experiments. In addition, conduct exploratory data analysis to understand the data further and select the relevant features that would assist in levelling the students across different education curricula.

Furthermore, to analyse the data, a machine learning (ML) based tool utilises cloud and fog paradigms to train the collected data and study the students' current/historical educational data to infer the best new level is designed and developed. The new tool allows stakeholders to monitor, adapt, enhance, and ensure the best usage of data to predict the correct status of students while switching between different education curricula. Test the proposed framework and viability of using these technologies (i.e., ML, Cloud and Fog computing) to solve the problem via several experiments. This involved seeking stakeholders' opinions on using the proposed tool for the quality assurance operation of transition between different education curricula. Finally, disseminate the results and findings in international specialised venues events and generalise the outcomes to other education curricula.

1.7 Thesis Structure

The remainder of this thesis is organised as follows:

Background and literature review (Chapter 2): This chapter discusses student levelling and the limitations and challenges in the UAE. Following that is the procedure currently used for student levelling and how each curriculum can differ in teaching style and students' levelling. This chapter discusses the literature review based on machine learning while discussing current ML models used to analyse student levelling and education-related datasets.

Machine learning and statistical tools (Chapter 3): This chapter discusses details about machine learning models, learning algorithms, and classification techniques. This chapter also discusses elements of the student levelling dataset while also providing an overview of the chapter.

Proposed methodology (Chapter 4): This chapter represents the proposed methodology framework and experimental setup for student levelling alongside ML classifiers to show its implementation. This chapter discusses the data preparation process like finding missing values and outliers, oversampling, and data normalisation. The experimental setup for the model is discussed in this chapter, with a summary of the methodology framework.

Results and Discussion (Chapter 5): This chapter discusses the results generated and analyses different machine learning models used in the simulations. This chapter discusses each classifier based on the performance evaluation methods used (Sensitivity, Specificity, precision, J1-score, F1-measure, accuracy, AUC, and ROC).

Web-Based Application System (Chapter 6): The following section will investigate the complete detailed information about the web-based student levelling system.

Conclusion and future work (Chapter 7): This section discusses the research outcomes and future work potential.

Chapter 2 Problem Background

2.1 Introduction

Chapter two discusses general information about student levelling in UAE's current procedures and methods used by schools to assess the level of students. The main objective of this research is to provide computational solutions for student levelling when transferring between curricula. A description of the current process used by the Ministry of Education (MOE) in UAE for student levelling is considered in this research, whilst also focusing on the process of admission and levelling followed by Belvedere British School in Abu Dhabi. Under the granted ethical approval for this research (UREC reference: 19/CMS/001), data for this research was collected from the Belvedere British school.

2.2 Overview of Education in Multicultural Countries

Transferring to a school in Abu Dhabi from another school in the UAE or transferring from an overseas school is undoubtedly more complicated than just completing some application forms. The Department of Education and Knowledge (ADEK) consistently seeks to deliver high-quality education and strives to ensure that children are assigned the correct age group and level when transferring [5]. Under the granted ethical approval for this research (UREC reference: 19/CMS/001), data for this research was collected from schools with different curricula. It can get more difficult for children to change the curriculum, mainly when each school uses a different grading structure.

2.2.1 School Admissions Process

ADEK implies strict rules when transferring between schools and curricula in Abu Dhabi. Before a student can be released from their previous school, parents must ensure a place in the new school is granted. Children are not permitted to transfer between schools or curricula within Abu Dhabi after 1st October for schools that follow a September to June academic year . ADEK has more flexibility for children transferring between schools of the same curriculum. However, schools sometimes apply their own rules concerning children who move from other curricula due to exam board requirements and subject compatibility.

Nevertheless, most schools in Abu Dhabi allow transfer between curricula at the beginning of the academic year; for British schools, it is up to Year Ten which is equivalent to Grade Nine in the American and MOE curricula. IGCSE is a two-year programme from Year Ten. Students who study this programme and are willing to transfer to another curriculum should do before this year. Students studying the American curriculum cannot move to another one after starting Grade Nine. This rule ensures that they complete all four critical stages to be eligible for the High School Diploma.

Since 2012, many students had repeated the same year when they transferred from grading to year system. Due to the large number of concerns raised by parents, ADEK developed a set of guidelines for schools and parents to follow and fulfil; some restrictions are shown**Error! Reference source not found.**. ADEK has created the policies that were explained to parents. In addition, they were made aware of the impact when moving from a grading system to a year system, and options were provided. Through consultation with schools, the best option is selected for their affected child from now on. When students transfer from the grading to the year system, they are expected to skip one year, e.g., moving from Grade One, they would go to Year Three. In some cases, the student has moved from Grade One to Year Two. ADEK has instructed schools in Abu Dhabi to identify the affected students and for consultations to be carried out with parents. Table 2.1 School transfer restrictions in Abu Dhabi according to ADEK rules.

Table 2.1 School transfer restrictions in Abu Dhabi

Age Group	Acceptance cut-off date	Conditions
FS1 to Year 6	All academic year	Approval of the new school
Year 7 to Year 10	1st week of term 2	Approval of the new school
Year 11 to Year 13	60 days from the beginning of the academic year	Approval of the new school and compliance with exam board regulations

Schools have invited parents whose child is affected by the transfer between the Grade and Year systems to discuss the student's status. Consultation with a parent includes reviewing any assessment undertaken by the student (e.g., CAT4, MAP, Arabic / Islamic test). Based on the assessment results and the social well-being of the child, the school will recommend an option for the parents on what year to assign their concerned child. If parents disagree with the school, they can transfer their child to another school. Figure 2.1 School admissions count for one school from 2018 to 2020 from different curricula.

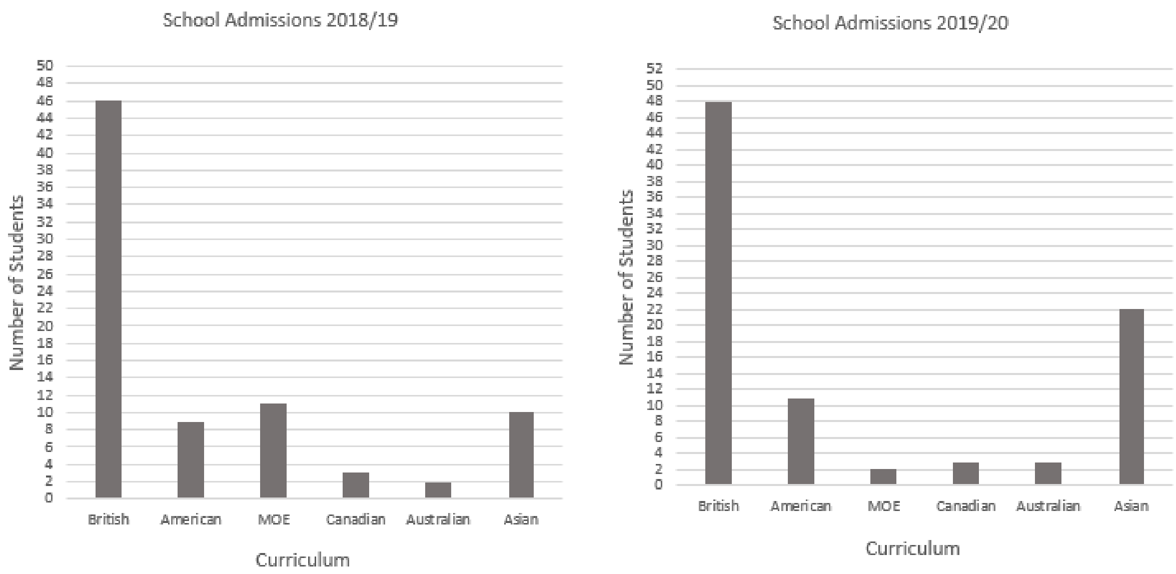


Figure 2.1 School admissions count for one school from 2018 to 2020

Currently, children who are transferring to the same curriculum need to start school in Abu Dhabi at least one month before any exams are taken so that the child is eligible to be promoted to the next year's group at the end of the academic year. In addition, schools in Abu Dhabi have the right not to accept children from another curriculum after the second term if they believe the transfer will negatively impact the child's learning. Although transfers after year ten tend to be difficult and restricted, there are certain conditions that ADEK has put in place to aid those willing to transfer in year ten. If the transfer happens in year ten, the new school needs to offer the same subjects as the previous schools and use the same exam board. There are cases where the exam boards are different. Then there is a three-way communication between parents, school and ADEK to ensure enough overlap between the two to make the transfer.

There is much emphasis on children who need to meet the mandatory age requirement for school entry, which is six years old on 31st December for schools following the September to June timetable. The school has year one as the appropriate year for children to start school. Most of the time, children start school at a younger age in their home country, which is sometimes more youthful than the standard criteria set by ADEK for the foundation stage and year one. Children that do not meet ADEK age criteria are most likely required to join the age-appropriate year group, even though they might have already studied at a higher year group in their home country. There are cases where children relocate from the Southern Hemisphere where their school year is the calendar year. Those children who join the year system where the school year starts in September can be required to repeat the same academic year since they haven't completed their school year in their home country. However, ADEK is flexible in this context. The school assesses the student based on that assessment alongside previous school assessments. An exception is sometimes made by the school allowing the student to join the next year's group. There are situations where sometimes a child transfers in time for term 2 (January) start and has completed the academic year before relocation. Then the child can join the next year's group.

Most schools in Abu Dhabi share similar admission processes. However, there are some variations that each school may implement to fulfil its admission process. Belvedere British School currently follows the admission process shown in Figure 2.2. According to this flowchart, the admission process for Belvedere British School has been designed specifically for this school. Yet, it is effective, but it is not consistent with other admission processes followed by different curricula.

Belvedere British School has two stages of the admission process, entry-stage, and the acceptance phase. At the entry stage, interested students submit their applications to the school. The admission officer will check previous records about the child and the primary school. If the documents are clear, the student can enter the school directly. They must pass an internal assessment with many criteria as shown in Table 2.2 or put in a holding pool depending on their availability. Once a student completes the internal review successfully, they will be transferred to the acceptance stage, where they are offered a place at the school. The acceptance procedure varies for students applying from overseas and at home. Students who come from overseas can join the school at any time. However, students who transfer from schools within the UAE can only participate during term registration, either in September and term 1 or 2.

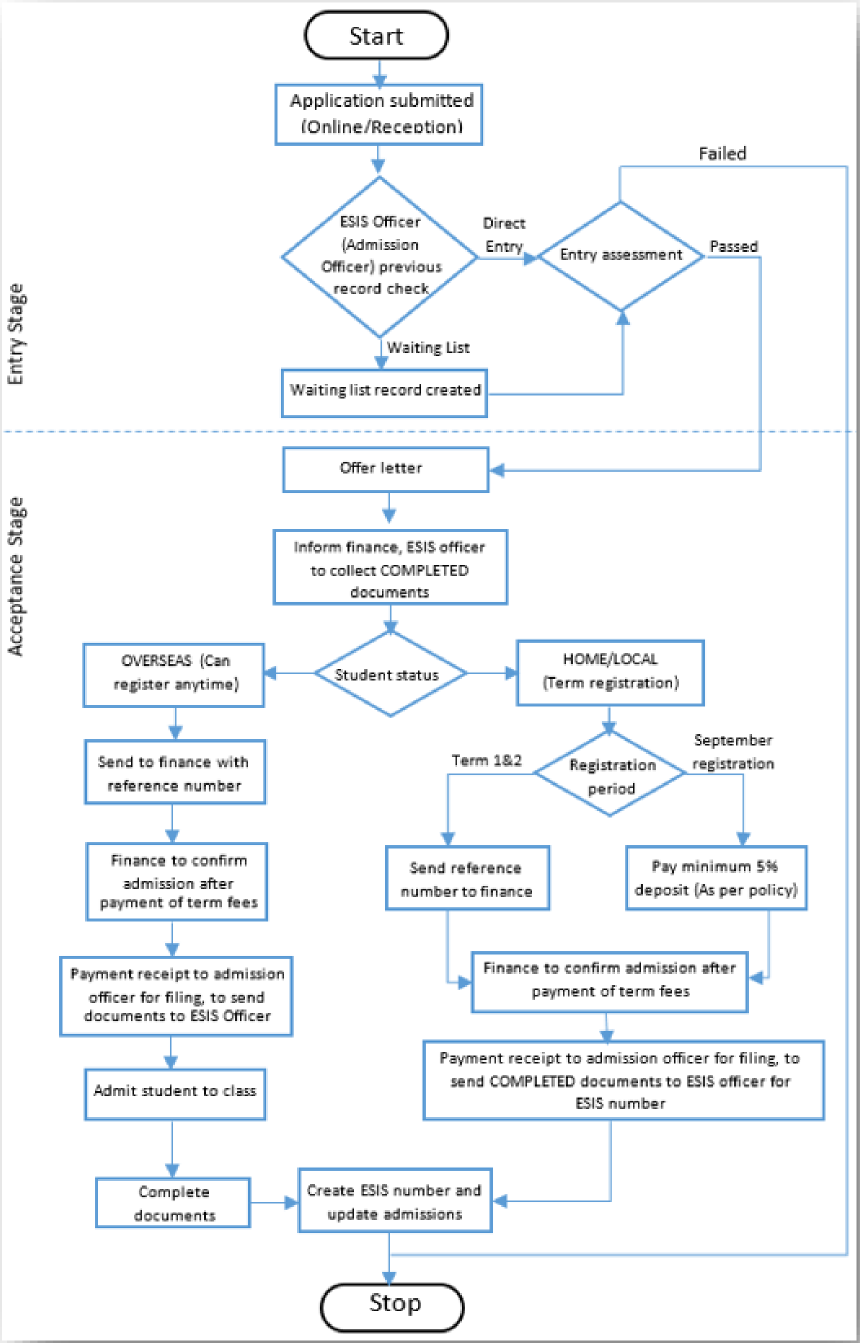


Figure 2.2 Standard admission process followed by Belvedere British School

Table 2.2 School admissions conditions and test plan

Year Group	Mathematics	English	Science	Conditions to meet
FS2	Interview Test	Interview Test	N/A	1 – Internal assessment
Year 1	30 Items	10 Marks – Writing and Comprehension	N/A	2 – Interview
Year 2	20 Items	10 Marks – Writing and Comprehension	N/A	3 – Payments
Year 3	30 Items	10 Marks – Writing and Comprehension	N/A	
Year 4	40 Items	10 Marks – Writing and Comprehension	N/A	1 – Internal assessment
Year 5	40 Items	10 Marks – Writing and Comprehension	N/A	2 – Interview
Year 6	40 Items	10 Marks – Writing and Comprehension	N/A	3 – Payments
Year 7	13 Items	25 Marks	21 Items	4 – Past education records
Year 8	19 Items	20 Marks	18 Items	5 – Attendance records
Year 9	36 Items	44 Marks	21 Items	6 – Discipline records
Year 10	49 Items	22 Marks	22 Items	

Belvedere British School currently bases its admissions on 50% test-based and 50% interview feedback. The acceptance or rejection of a student into the school is based on the below conditions:

Accepted – Successfully passed both admission test and interview or did not pass admission test but reconsidered after the interview feedback with conditions set.

Declined – Did not pass the admission test twice or may have passed but failed the interview.

The school has several students who had moved from a Grade system to a Year system and repeated a year when assigned to their year group. Parents have raised those concerns to the school to solve the child. Therefore, ADEK and the school have provided parents with three options during admission, as described below:

For students in FS to Year 8 only, the parents were given the below options to resolve the discrepancy between their age and assigned year group:

Option 1: *Students will continue in their current path and will be promoted to the next year group (e.g., a student who has completed Year 3 will be moved to Year 4)*

Option 2: *Students will be moved to a higher year by two-year groups and offered academic support "Bridging" program at the beginning of the academic year (e.g., a student who has completed Year 3 will be moved to Year 5).*

Option 3: *Students will be transferred to a grading system and moved to the next grade.*

The school has the authority to decide which option to place the child based on some factors such as academic capability, social wellbeing, and the school's operational constraints. Parents always have the right to choose option three if they cannot go to a final agreement with the school.

Students in Years 9 and 10:

Option A: *Will continue in their current path and will be promoted to the next year's level*

Option B: *Will transfer to a grading system and will be moved to a higher-grade level*

2.2.2 Student Levelling Process

The students' levelling is based on cognitive abilities testing, which has different levels as shown in Table 2.3 and usually takes place during September of each academic year to precisely determine each child's level and placement. The tests are either online or paper-based and have two versions, group reports and individual reports, which are then shared with the teachers to identify the student's actual level in the class.

The online levelling is done using an external agency that generates a rounded profile of the child's ability so the school can target areas where support is needed and provide the right level for the child to make informed decisions about their progress. The assessment provides a profile of strengths and weaknesses in four subject areas: Verbal Reasoning, Numerical Reasoning, Non-verbal Reasoning, and Spatial (UREC reference: 19/CMS/001).

Table 2.3 External agency assessment level used for each year's group

Assessment Level	Year Group
Level X	Year 2
Level Y	Year 3
Level A	Year 4
Level B	Year 5
Level C	Year 6
Level D	Year 7
Level E	Year 8
Level F	Year 9 & 10
Level G	Year 11 & 12

2.2.3 Individual Student Differentiation

Every student has a report generated once a year that indicates the student's strengths and weaknesses based on the scores obtained at the beginning of the year. As shown in Figure 2.3, students are placed to their proper level based on the age group ranging from 1 to 9 and C, B, and A for each class.

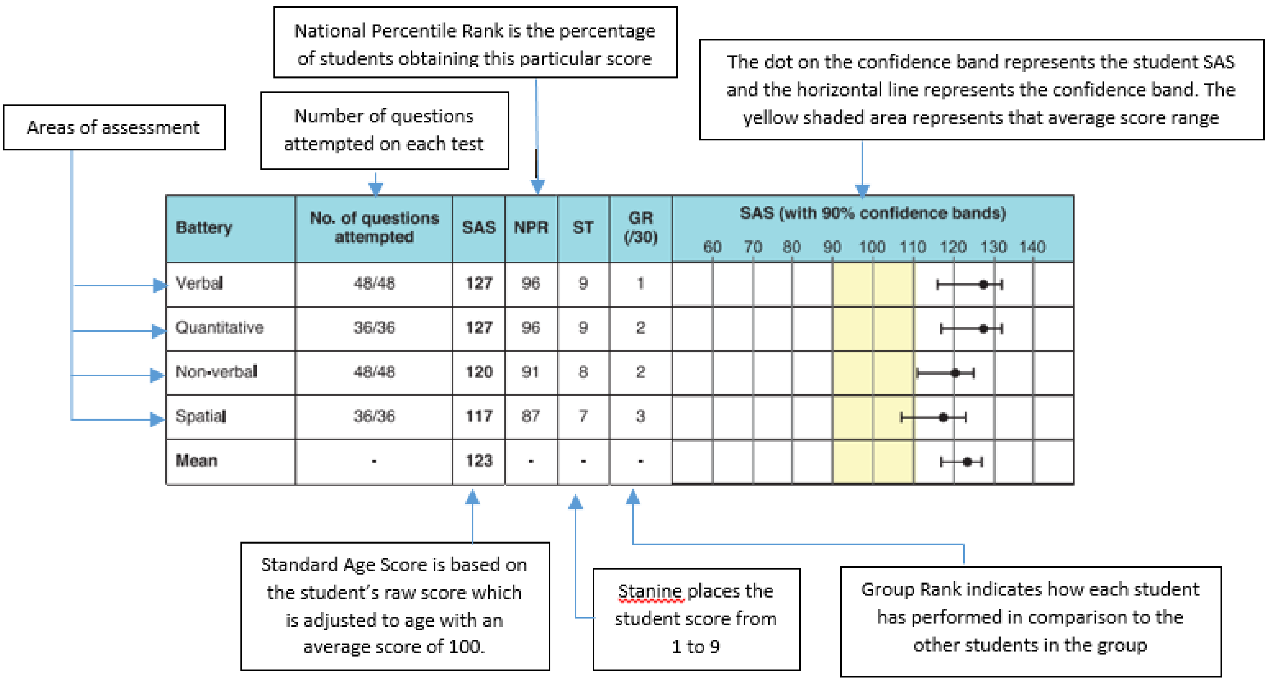


Figure 2.3 Student report generated at the start of the academic year

2.3 Summary

The education sector underutilises advanced technologies and cannot proficiently increase operational efficiency. Most of the students' records are on traditional data-saving programmes such as excel. There are gaps identified in school operating procedures in admissions, levelling, and differentiation. The admissions procedures currently used in schools across Abu Dhabi are inefficient in time consumed, and errors caused. In general, the current admission process takes approximately one week for an individual student. It has been discovered that student levelling is not consistent throughout the schools because some schools do their levelling while others use external agencies to do their levelling.

Besides, some schools do not provide levelling for their students until the end of the year. Therefore, schools do not have a standard procedure amongst them. Finally, differentiation is a form of levelling that is hardly found in schools in Abu Dhabi. Every student shares the same work in class and at home, and there was no guide to identifying the weak and the strong students. As a result, some students are currently in the wrong year/grade group, and students that are weak in certain subjects are given the same work as the stronger students, which creates a significant gap in their learning. This framework will assist schools in saving complete information about the students, analysing the student background of education, proposing their eligibility, predicting the student performance, and being a rebus platform for developing the possibilities of adding new modules to the system.

A cloud-based application can help share information, allowing everyone to store data of interest in the same place, which would permit the school to update students' information easily, input their progress examinations, and share the data with other stakeholders to facilitate decision-making. It can support datasets, including various excel sheets, graphs, and images. Unlike the contemporary methods, it enables on-demand access to significant storage and computing facilities. Given that the cloud applications for the education sectors require a high level of security, availability, and privacy, private or hybrid clouds come as a handy option to act as the root for its implementation.

The proposed system considers admissions, levelling, and differentiation for schools. With the proposed plan, entries will be consistent between all schools as they use the same system and are time efficient. Current systems do not allow schools to discover the level of the student immediately. They can only rely on previous school reports and entry exams if provided. Still, the proposed system can generate the year/grade group, level of the student, and if any differentiation is required. Student academic information can be viewed anytime and anywhere to allow teachers to assess their students regularly.

Chapter 3 Literature Review of Machine Learning and Statistical Tools

3.1 Introduction

This chapter represents the literature review of machine learning algorithms and statistical tools implemented in this research. It elaborates in detail on the learning algorithm types, such as supervised and unsupervised learning techniques, which are considered the most important domain in machine learning. The second section of this chapter considers machine learning algorithms and extracting useful information from student-levelling data, including classifying the data. Nevertheless, the data selection criteria are discussed, concentrating on the student-levelling dataset. The last section of the chapter presents statistical tools and techniques, while a conclusion is provided at the end of the chapter to summarise the discussed topics.

3.2 Machine Learning Algorithms Descriptions

Machine learning algorithms are based on Artificial Intelligence, allowing an operating system to undergo learning without the need for explicit programming [4], [5]. This kind of AI method can be implemented using numeric datasets jested into trained ML models to solve challenges related to numeric predictions, pattern recognition and classification [6]. The machine learning classification process starts with a training set containing known target value instances. Subsequently, a testing set is used that has unknown samples. The classification performance can be evaluated based on the test count instances that the model predicted correctly and incorrectly [7].

Two main techniques are used in machine learning: Supervised learning and Unsupervised learning. Because the dataset I have collected is not big enough to apply to Unsupervised learning, I used the Supervised learning technique.

3.2.1 Supervised learning technique

Most practical machine learning uses supervised learning. Supervised learning is used in which an input variable (x) and output variable (y) are to learn the mapping function from the input to the output [8]. The purpose of supervised learning is to approximate the mapping function close to accuracy so that when a new input data (x) is introduced, the output variable (y) can predict and generate instances for the new data inserted. Supervised learning is anticipated to find the data patterns applied to an analytics process [9]. The main aim of the training set is to learn from labelled instances to identify unlabelled ones during the testing phase with high potential accuracy.

The training process in supervised learning continues until the algorithm achieves acceptable accuracy on the training data. The desired output must be known and considered a typical relationship between the input and output value [10]. For example, a training set has students with different average test marks (50%, 70%, and 80%), and the learner is given the level of the student based on past exam marks. The test set holds students with an unknown class label to identify that label. In the training stage, the class label is specified to the classifier. The supervised learning process operates with known inputs combined with known outputs[11].

In this research, I used the supervised learning technique by calculating the average of the last year's final marks and dividing them into three classes (class 1, class 2, and class 3), which class 1 is from 85% to 100%, class 2 is from 75% to 84.99%, and class 3 is below the 75%. These classes will be the label of the dataset for applying the supervised classification algorithms.

3.3 Literature review on the use of Machine Learning in Education

In recent years, education sectors have faced many challenges to meet the demands of e-Learning. The primary motivation for researchers is to support education institutes in terms of student assessment, learning behaviour and student transition between schools [12]–[14]. Some machine learning studies are conducted on student levelling, which are related to this research. In this section, the influence gained from different studies and their limitations for the analysis of student levelling will be presented.

Masci et al. [15] developed and applied novel machine learning and statistical methods to analyse why students' PISA (Program for International Student Assessment) 2015 test scores in nine countries: Australia, Canada, France, Germany, Italy, Japan, Spain, UK, and the USA. The study aimed to determine which student characteristics are associated with test scores and which are associated with school value-added (measured at the school level). The researchers applied a two-stage methodology using flexible tree-based methods to address these issues; first, they ran multi-level regression trees to estimate school value-added. The second stage relates the estimated school value-added to school-level variables using regression trees and boosting. Results show that while several student and school-level characteristics are significantly associated with students' achievements, there are marked differences across countries. The proposed approach allows an improved description of the structurally different educational production functions across countries. Their study has focused on ways the school has impacted student performance; however, it doesn't discuss the student's actual performance just based on their input data portions.

Shabandar et al. [16] researched Machine Learning approaches to predict learning outcomes in Massive Open Online Courses (MOOCs). Research is available within the area of MOOC data analysis, considering the behavioural patterns of users. Based on learner behavioural patterns, two sets of features were compared in terms of their suitability for predicting the course outcome of learners participating in MOOCs. Various machine learning algorithms have been applied to enhance the accuracy of classifier models. Simulation results from the investigation showed that Random Forest achieved viable performance for the prediction problem, obtaining the highest version of the models tested. The study has some relation to this study focusing on student performance.

Hsia et al. [17] conducted a study using data mining techniques to analyse course preference and the completion rates of enrolees in extension education courses (after graduation courses) in a university in Taiwan. Using their dataset, the researcher aimed to improve the target curriculum based on the student's needs. The researcher used Decision Tree Algorithms, Link Analysis Algorithms, and Decision Forest Algorithms. The data collected from the university was in a range of five academic years from 2000 to 2005. Overall, 1408 records were compiled. After testing eight different algorithms and studying their capabilities in classification, prediction, clustering and description, the researcher selected Decision tree, Link Analysis and Decision Forest as the best algorithms for this study. Three separate variables were chosen, taking into consideration the desired outcome of the study. The variables selected were course category, completion status, and enrolee profession. Based on the results gained from the study, the Extension Education Centre at Chienkuo Technology University can plan for future courses based on the needs of the students. Masci's study is closely related to this thesis since the researchers have used several features related to student needs, such as student learning outcomes and their capabilities in different subject areas. Hence, further testing could have been implemented to show if a student were suitable for a specific course or not based on past collected data.

Lykourentzou et al.[18] researched the dropout prediction method for e-learning courses based on three popular machine learning techniques and detailed student data. The machine learning techniques used in the study are feed-forward neural networks, support vector machines, and probabilistic ensemble simplified fuzzy ARTMAP (Adaptive Resonance Theory). Since using a single algorithm may not provide accurate results, three different machine learning techniques were used to predict student dropout for e-learning courses. The method was examined in terms of overall accuracy, sensitivity, and precision. Its results were significantly better than those reported in the relevant literature. The research provides vital information for the education sector, the dropout prediction. However, they focused on e-learning courses rather than face-to-face education in universities. Student-level can be predicted based on past and current exam marks in school using the method conducted in this paper. The student's performance can be expected to be successful or not in courses based on past performance using those prediction models.

3.4 Classification

This research uses supervised classification because the dataset collected from the schools has been identified with relevant labels. Regression models are aimed at mapping instance (input) values to continuous outcome values, while classification procedures aim to map samples (input) into discrete classes [19]. For example, some studies aim to classify students that will drop out of their exams or otherwise they will not. Within classification, the aim is to learn a decision-making platform that can correctly map an instance (input) space to an output. Classifying student and education data has shown positive outcomes for student benefits and education institutes [20].

In the classification process, consider the object (x) as the input with a set of features, while (y) as the class label assigned alongside x. The classification model is implemented to predict the class label for new samples. Classification techniques in education are essential to improve student levelling and enhance education institute decision-making in terms of grading and student abilities [21]. Various methods are implemented for classification, grouped into two linear and nonlinear classifiers. As represented in Equation 3.1$g(x) = w^T x + bg(x) =$

$w^T x + b$, the linear classifier is described as a linear function (g) of the input (x), (w) represents a set of weights, and (b) refers to the bias.

$$g(x) = w^T x + b \qquad (3.1)$$

Nonlinear classifiers discover the class of a feature vector x using a nonlinear mapping function *(f),* where the function *f* learns from a training set T. Accordingly, the model develops the mapping required to predict the new data correctly. The most used nonlinear classifier is Artificial Neural Network (ANN). ANN has a couple of output units based on each class [22]. ANN is manifested with a direct connection of weights connected through nonlinear transfer functions. The consequences must be configured accordingly for the ANN model to perform the necessary tasks to receive valid results from the developed learning algorithm. During the learning phase, use an optimisation algorithm to find ways to reduce an error originating from the objective function of interest. The dimension of variation influences the effectiveness of the NN, including things like network connectivity pattern, activation functions, determining the appropriate weights, and finally, the training data ingested into the model during the learning phase. The computation at a single node is the weighted sum of the input, and then, as a result, it's calculated based on the activation input. An example of this computation is shown in Equation 3.2, where y_j is the output from the j^{th} unit in layer y, w_{ji} represents the weight of the i^{th} input, x_i represents the value of the i^{th} input, and σ represents the activation function.

$$y_j = \sigma\left(\sum_{i=0}^{m} w_{ji} x_i\right) \qquad (3.2)$$

3.5 Student Levelling Dataset

The UAE is a multicultural country with many different nationalities relocating from worldwide. To meet expatriates' needs, United Arab Emirates has established many international private schools. However, since every country has a distinct curriculum, many challenges were faced by schools and the MOE in allocating students to their correct year/grade group and keeping track of their academic performance while moving between the curricula and assigning students to a proper level. Consequently, these data are essential to show students' levelling in multiple curricula. Also, these data help highlight how students' progress can vary when they transfer between curricula [23]. In this research, the collected dataset from the schools comprises novel aspects specifically in terms of student grading in diverse educational systems within the UAE – Researchers and other education sectors can use this data to see the impact of having varied curricula in a country. The dataset can be used by intelligent algorithms, specifically machine learning and pattern analysis methods, to develop an intellectual framework applicable to multicultural educational systems and aid in a smooth transition by minimising the impact of switching the students' academic curriculum [16], [24]. Table 3.1 shows the description of the raw dataset attributes.

Table 3.1 Student levelling dataset description

#	Attribute Name	Value	Description
1	Student Name	First and Family Names	Full Name of students
2	Student ID	Number	ID number of students "within the school."
3	Gender	Male/Female	Gender of the student
4	Date of Birth	Date	Student's Date of birth
5	Proposed Year/Grade	Year 1, 2, 3, etc. / Grade 1, 2, 3, etc.	The year or grade group assigned to the student by the school
6	Year of Admission	2017-18 /2018-19	The collected data for two academic years; 2017/2018 and 2018/2019
7	Previous Curriculum	The UK, US, UAE, Canadian, Indian etc.	The curriculum that the student transferred from
8	Current Curriculum	British / American	The curriculum of the student transferred to
9	Previous Year/Grade	Year 1, 2, 3, etc. / Grade 1, 2, 3, etc.	The year or grade the student was assigned to in their previous school
10	Math Entry Exam Mark	Mark out of 40	Exam marks for school entry exam in Math
11	Science Entry Exam Mark	Mark out of 40	Exam marks for school entry exam Science
12	English Entry Exam Mark	Mark out of 40	Exam marks for school entry exam English
13	Maths Marks 19-1	Percentage out of 100%	Term 1 student Maths Exam marks during the academic year 2018/19
14	Science Marks 19-1	Percentage out of 100%	Term 1 student science Exam marks during the academic year 2018/19
15	English Marks 19-1	Percentage out of 100%	Term 1 student English Exam marks during the academic year 2018/19
16	Maths Marks 19-2	Percentage out of 100%	Term 2 student Maths Exam marks during the academic year 2018/19
17	Science Marks 19-2	Percentage out of 100%	Term 2 student science Exam marks during the academic year 2018/19
18	English Marks 19-2	Percentage out of 100%	Term 2 student English Exam marks during the academic year 2018/19
19	Maths Marks 19-3	Percentage out of 100%	Term 3 student Maths Exam marks during the academic year 2018/19
20	Science Marks 19-3	Percentage out of 100%	Term 3 student science Exam marks during the academic year 2018/19
21	English Marks 19-3	Percentage out of 100%	Term 3 student English Exam marks during the academic year 2018/19
22	Maths Marks 20-1	Percentage out of 100%	Term 1 student Maths Exam marks during the academic year 2019/20
23	Science Marks 20-1	Percentage out of 100%	Term 1 student science Exam marks during the academic year 2019/20
24	English Marks 20-1	Percentage out of 100%	Term 1 student English Exam marks during the academic year 2019/20
25	Maths Marks 20-2	Percentage out of 100%	Term 2 student Maths Exam marks during the academic year 2019/20
26	Science Marks 20-2	Percentage out of 100%	Term 2 student science Exam marks during the academic year 2019/20
27	English Marks 20-2	Percentage out of 100%	Term 2 student English Exam marks during the academic year 2019/20
28	Maths Marks 20-3	Percentage out of 100%	Term 3 student Maths Exam marks during the academic year 2019/20
29	Science Marks 20-3	Percentage out of 100%	Term 3 student science Exam marks during the academic year 2019/20
30	English Marks 20-3	Percentage out of 100%	Term 3 student English Exam marks during the academic year 2019/20

3.5.1 Multi-Class classification

The multi-class classification problem refers to assigning each observation into k classes. Two-class issues are found to be easier to solve. Therefore, many researchers use two-class classifiers [25]. However, many researchers have examined the performance of learning models with more than the two classes classification dataset [26]. Such classifiers perform effectively to compare the probability of different labels and distinguish them with the highest probability. The proposed study focuses on the relative proportion of several errors such as sensitivity and specificity to check the student level using the classification method.

3.6 Statistical Tool Selection

3.6.1 Feature Selection

Feature selection provides an effective way to develop prediction performance, reduce computation time, and provide a better understanding of the student-levelling dataset in ML models.

Akour et al. [27] stated that working with ML tasks requires a more efficient technique to provide the needed outcome due to the increase in data dimensions. In recent years, many researchers have proposed numerous methods and techniques to reduce high data dimensions and receive the desired accuracy. Dimensionality reduction is used to increase the accuracy of learning features and decrease the training time as a pre-processing step. This pre-processing step can eliminate irrelevant data, noise, and redundant features. DR is performed based on FS and FE [28]. Using FS is vital because data is generated continuously at an increasing rate. Therefore, dimensionality problems can be reduced by decreasing redundancy and eliminating irrelevant data [29]. On the other hand, FE finds the most distinctive, informative, and decreased set of features to increase the efficiency of the processing of data whilst increasing the efficiency of the storage of data [30].

Ramaswami et al. [31] researched feature selection techniques in educational data mining. The study aimed to discover the effectiveness of the student performance model in connection with the feature selection technique. In this case, six filtered feature selection algorithms were used, and the F-measure and ROC were utilised as quality measures. As a result, the outcome showed a reduction in computational time and constructional cost in the student performance model's training and classification phases[31].

Sivakumar et al. [32] collected a dataset of 240 students using a survey in a university located in India, containing thirty-two features of their socio-demographic, academic and institutional information and applied a Correlation-Based Feature Selection (CFS) algorithm in the pre-processing step. Consequently, the model's accuracy showed more than 90% using one dataset.

On the other hand, Zaffar et al. [33] conducted various feature selection algorithms and analysed their performance using two different datasets. Their results showed a significant performance difference using a different number of dataset attributes, which led to a 10% to 20% change in the accuracy. They discovered that the effectiveness of the feature selection techniques decreases as the number of features grows.

There are two procedures followed when selecting the correct feature subsets. Firstly, finding the possible feature subsets is required. After that, the feature subset is determined based on the objective function. Once those two steps are completed, the feature subsets can be implemented and used by ML algorithms. Figure 3.1 describes the basic flow of the feature selection methodology used in this research.

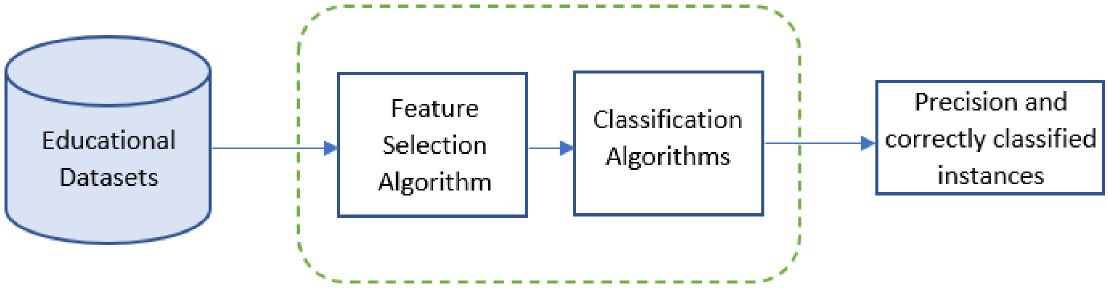


Figure 3.1 Flow of Feature Selection Methodology

3.7 Decision Tree Algorithm

The decision tree algorithm belongs to supervised learning algorithms. A decision tree algorithm can solve regression and classification problems [34]. It can be used to create a training model that can predict the class or value of the target variable by learning simplified decision rules from prior training data. The decision tree algorithm begins from the tree's root to predict a class label for a specific record [35]. The values of the root attributes are compared with the characteristics of the record and then follow the branch corresponding to that value and then move on to the next node [36].

Two decision trees are based on the target value [37]:

1. Categorical Variable Decision Tree: A decision tree with a definite target variable.
2. Continuous Variable Decision Tree: A decision tree with a constant target variable.

Decision tree models classify the examples by guiding them through the tree starting from the root to the leaf/terminal node, where the leaf/terminal node provides the model's classification. Every node in the tree acts as a test case from some attribute, whilst every edge descending from the node is every possible answer to the test case [37]. The whole process is repeated when reaching every subtree rooted down to the next node. Classification and regression trees both have different decision criteria [38]. Decision trees use multiple algorithms to divide a node into two or more sub-nodes. They follow the Sum of Product (SOP) representation, and the SOP is also known as a disjunctive standard form [37]. For every class, the branches from the root to the leaf node of the tree having the same category are conjunction (product) of values, whilst different components ending in that class form a disjunction (sum) [37].

The target variables influence the algorithm selection. Several algorithms can be used in decision trees [37]:

- ID3 – Extension of D3
- C4.5 – Successor of ID3
- CART – Classification and Regression Tree
- CHAID – Chi-square automatic interaction detection works on multi-level splits when computing classification trees
- MARS – Multivariate adaptive regression splines

Several algorithms can be used when building a decision tree. The most popular and heavily implemented method is ID3 and C4.5 developed which a successor was developed by Quinlan [39]. There are standard components that decision tree algorithms have. The ID3 algorithm is built using a top-down greedy search method through all the possible branches in the model and without backtracking [34]. A top-down greedy search constantly searches for the best choice at that moment. There are several steps to be followed when using the ID3 algorithm [40]. The algorithm begins with the original set S at the root node. On every iteration of the algorithm, the Entropy (H) and Information Gain (IG) of the attribute are calculated based on the iteration through unused features of the set (S). The attribute with the lowest (H) or the largest (IG) is selected whilst the group (S) is then split by the selected feature to develop a subset of the data [41]. The algorithm is repeated for each subset but only considers previously not chosen attributes.

The purpose of having an effective decision tree is to have the ability to generate the most information gained from different features. IG is calculated based on Equation 3.3 [42].

$$Information\ Gain = entropy\ (x) - ([\ the\ Avarage\ weight] \times entropy\ (feature))$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X) \qquad (3.3)$$

There are several steps to be followed to calculate the information gain. In the first step, the entropy of the target values must be calculated, and then the data used for the model is split into different features. The entropy method is implemented when calculating all the attributes. However, the total entropy is subtracted before splitting the dataset into branches. Finally, the samples are divided based on the maximum IG. Algorithm 3.1 represents the process of breaking data in a decision tree [43].

Algorithm 3.1: Building a Regression Tree

1. It is required to use recursive binary splitting to build large trees using the training dataset, stopping when a terminal node has fewer than the minimum number of observations.
2. Employ cost-complexity pruning to the large tree to achieve a sequence of the optimal subtree as a function of α.
3. Apply K-fold cross-validation to choose α and divide the training set into K folds. For each $k = 1\ldots, K$:
   (a) Repeat Steps One and Two on a *k*th fold of the training sets.
   (b) Evaluate the error rate using the mean squared prediction on the testing sets (left out of *k*th fold) as a function of α.
   (c) Calculate the outcomes by averaging each value of α.
4. Return the subtree from Step 2 for choosing the value of α as corresponds to that.

In this thesis, a decision node (level of the student) has three branches (Class 1, Class 2, and Class 3). The nodes denote a classification target value or a decision, and the root node provides the optimal predictor. Figure 3.2 represents an example of a decision tree that can be implemented.

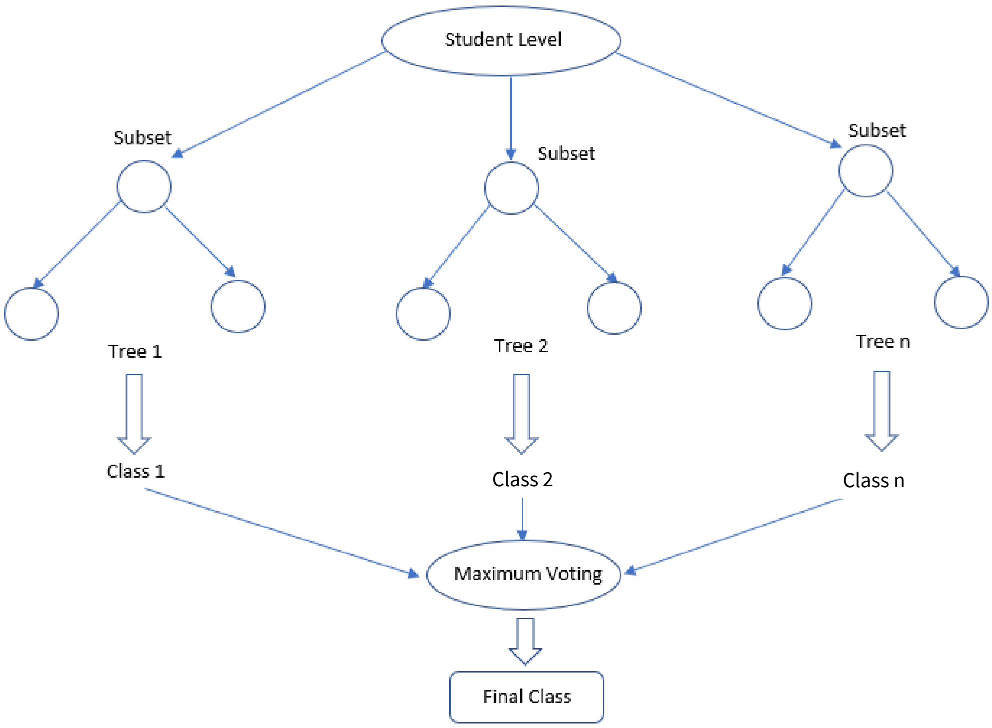


Figure 3.2 Decision tree example [44]

3.8 Random Forest Classifier

Random forest classifier is one of the most successful ensemble learning techniques proven to be effective in pattern recognition and ML for high-dimensional classification and skewed problems [45]. The RFC method was initially introduced by Kam Ho in 1995 [46] and then further developed by Breiman [47]. In practice, it is common for a slight change in the training dataset to be in a different tree, which is the hierarchical nature of the tree classifiers [48].

For the last two decades, random forest classifiers have received much attention due to the effective classification results and their speed of processing the data [49], [50]. The RFC yields reliable classification using predictions from various decision trees [47].

Decision trees are constructed based on many features; those features are selected randomly at each tree node [51]. The attributes relevant for classification are found by calculating the importance score for each element [51]. The decision trees are randomised using a bootstrap statistical resampling technique alongside random feature selection [51]. The optimum split is calculated utilising the *m* features until the tree has developed without pruning. The process is repeated for all the trees in the forest (refer to Algorithm 3.2) using different bootstrap samples of the data [51]. After that, the classification of new instances can be gained using a majority vote, and this procedure uses bagging combined with decision tree classifiers to achieve this. When building the decision trees, a split is required from the complete set of p predictors in the tree for every random instance of m predictors [43].

Algorithm 3.2: Random Forest

1 Given a training set $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $x_i \in R^d$ and $y_i \in C$, where C represents target classes; define the B of trees and the m of random features to select.

2 For b = 1, ..., B,

3 Using the training set of dataset and sampling produces a bootstrap instance of size n; some patterns in the training set will be replicated again, while others will be omitted based on the tree itself.

4 Implement a decision tree model, $\eta_b(x)$. For utilising the bootstrap example as a training dataset, each node in the tree m variables is randomly selected for splitting.

5 Classify the out-of-bag data (the non-bootstrap patterns) using the $\eta_b(x)$ model.

6 Assign x_i to the target class most characterised by the $\eta_{b'}(x)$ models, where b' belongs to the bootstrap instances that do not involve x_i.

This method can generate a couple of trees to form a big forest. The more trees in the forest, the more robust the algorithm generates high accuracy [52]. A substantial increase empirically and theoretically from the development of decision tree ensembles can be formed to produce a final decision using a voting procedure. To develop such ensembles, RFC follows the process of feature bagging [47]. With ensemble design, weak RFC decision trees can be improved to become stronger learners [52]. The model can use the empirical dataset to train and test them efficiently to predict confidence and estimate test errors.

Spoon et al. [53] explored multiple ways to utilise random forests in a learning analytics setting, emphasising approaches to identifying at-risk students and determining how to characterise students who would benefit the most from a particular intervention. As an illustration of the ability of a random forest to predict a continuous variable, the random forest for the 960 students in the final exam outcome data set had a resulting out-of-bag Mean squared error (MSE) of 2025.95. While these predictions were better using training and testing datasets with linear regression, advising students to enrol in Stat 119A based on low predicted final exam scores, which account for only 30% of the overall grade, is not as straightforward as using the successful completion outcome.

Cardona et al. [54] applied RFC to predict student degree completion rate. To improve retention rates, the research suggested that colleges and universities require strategies for intentional advising to ensure that students can complete their majors promptly. Currently, efforts have been made to adjust admission requirements; however, retention rates are still low. These strategies have reduced access to higher education for students from different economic sectors. Thus, institutions have recognised the need to understand better the factors that impact retention to focus their efforts. To this end, this research presents the application of random forests to predict degree completion within three years, representing 150 per cent time to completion, and identifies the variables that impact student retention at a large community college in the Midwest. The random forest algorithm consists of bagging (combining) decision trees created randomly from the training sample, thus creating a "forest". The model in the study was developed using data on 282 students with 14 variables. The variables included student details such as age, gender, degree, and college GPA. The model results, which have prediction and variable ranking, offer a critical understanding of developing a more efficient and responsive system to support students.

3.9 Artificial Neural Network

Artificial Neural Network (ANN) is a topic that has been researched since the 1980s. The method of ANN abstracts the human brain neural network from the way it processes information while also determining a model and establishing different networks according to other connections [55].

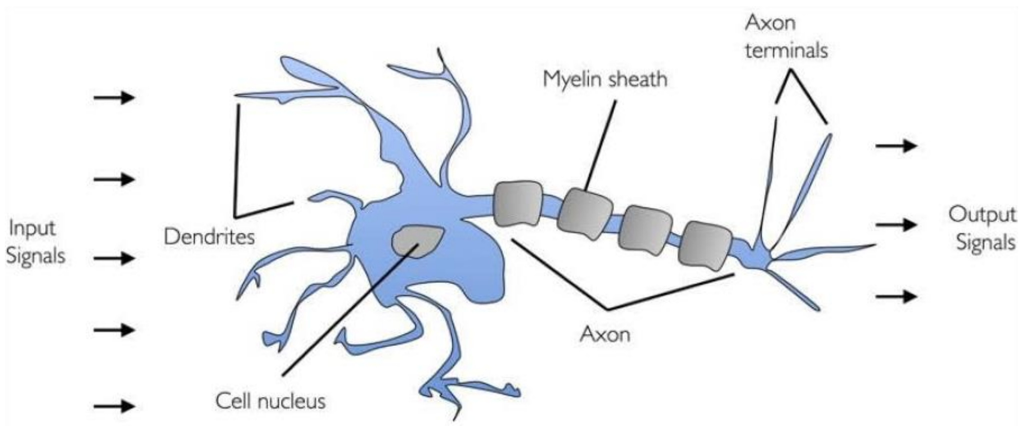


Figure 3.3 Example of a biological Neuron [56]

As shown in Figure 3.3, the Neural Network is a computing model which has many nodes (or neurons) connected [57]. Each node in the model represents a specific output function referred to as the activation function. Weights represent the interconnection between two nodes for the signal that passes through the connection between the two nodes. The weights are also equivalent to the memory of the ANN [58]. The network's output varies depending on connecting the network, the weight value, and the incentive function.

In ANN, a neuron processing unit can represent features, letters, or different concepts. The type of processing method in the ANN model is divided into three distinct sections: input unit, hidden unit, and output unit. The input unit of the network accepts signals and data from outside [59]. The hidden unit is located between the input and output units. The output unit generates the output of the system processing results [60]. The weights connected between the neurons reflect the strength between the cells. The representation and processing of information are embodied in the connection relationship of the network processing unit. ANN is a parallel distributed system that follows a different approach than traditional AI and information processing technologies [61]. ANN overcomes the defects of the conventional logic-based artificial intelligence in handling intuition and unstructured information and has the advantages of adaptive, self-organising and real-time learning features [61].

ANN has been widely used to study the behaviour and control of animals and machines. However, ANN has also been used for pattern recognition, forecasting, and data compression [62]. ANN consists of many layers; input layer, hidden layers (one or more according to the need), and output layer. The inputs (like synapses) are multiplied by weights. Weights assigned with each arrow represent information flow [62]. These weights are then computed by a mathematical function that determines the neuron's activation [63]. Another function computes the output of the artificial neuron [63]. The neurons of this network sum their inputs. Since the input neurons have only one, their output will be the input they received multiplied by a weight, as shown in Figure 3.4 [62]. If the weight is high, then the input will be substantial. By adjusting the weights of an artificial neuron, the output can be obtained for specific inputs. Algorithms can be found to adjust the weights of the ANN to get the desired output from the network. This process of adjusting weights is called learning or training. The training begins with random weights, and the goal is to change them so that the error will be minimal [62].

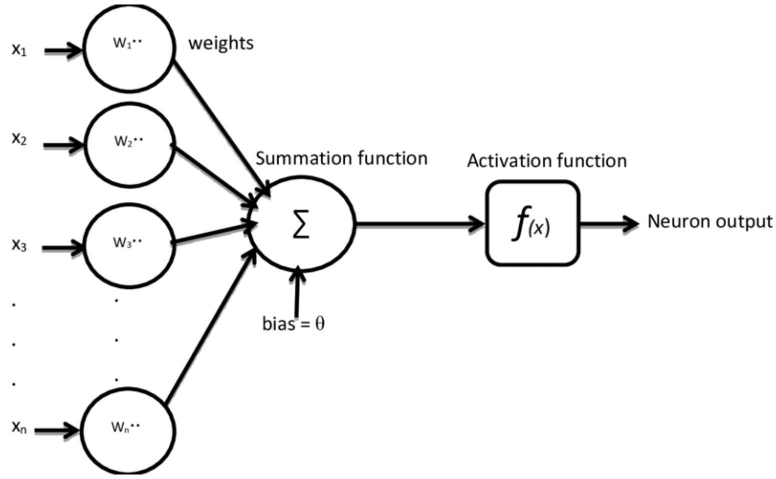


Figure 3.4 Node of the Neural Network [64]

3.9.1 Backpropagation Network Classifier (BPXNC)

Most variations of the Neural Network Algorithms derive from the four-layer backpropagation neural network [65]. The most used input/output configuration is one input node for every input channel and one output node for every class label. For the training stage of supervised learning, the network weights are adjusted in an iterative, gradient descent training procedure called backpropagation [66], [67]. The training data consists of a pair of data vectors. The input data is the pattern that will be learned, whilst the output data is the desired set of values. The main aim of the training is to reduce the overall error produced between the desired output and the actual output of the network. To guarantee a decrease in error, the incremental adjustments in the weights at each iteration must be tiny. A learning rate parameter must be specified for the network to improve the training time. The learning rate parameter is the percentage of the step taken towards minimum error. If this quantity is too small, training will take too long, and if it is too large, the gradient descent will degenerate, and the error will increase. Backpropagation is not always guaranteed to find the global minimum error. The NN takes the gradient descent from the current position to one with a lower error [68].

There is a possibility that the system may oscillate between two points. If the network reaches a local minimum in the error space, there is a possibility that it can be stuck, and the error will not reduce. Lippmann [69] discussed the dependence of the decision regions on the number of network layers and nodes per layer. He trained the network results to form the decision boundaries in the feature space. He showed that a three-layer configuration could generate any convex region for a network with threshold activation functions in the feature space. Figure 3.5 represents the learning process of the backpropagation algorithm [70].

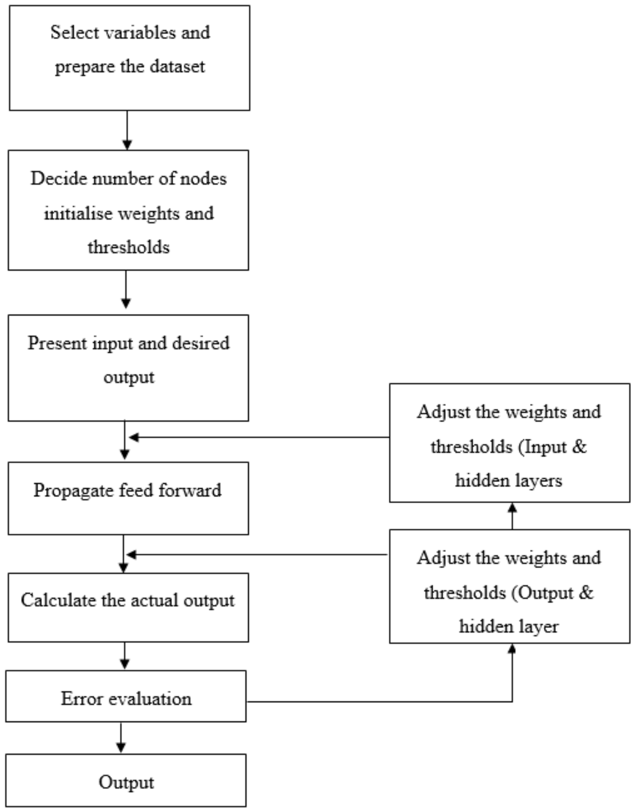


Figure 3.5 Flow chart of the Backpropagation Algorithm

3.9.2 Levenberg Neural Network (LEVNN)

Levenberg-Marquardt Algorithm is usually used as a standard algorithm for training the Neural Network to solve nonlinear least-squares problems. A combination of gradient descent and Gauss-Newton methods appears in this algorithm. In many cases, the LEVNN can guarantee problem-solving through its adaptive behaviour [71].

The Gradient Descent with adaptive learning rate Algorithm (GDAs) updates the weights and biases in the direction of the negative gradient of the performance function. Unlike GDAs, Conjugate Gradient Algorithm (CGAs) searches for the steepest descent and conjugate directions. Quasi-Newton Algorithm (QNAs) converge faster than CGAs and give better-generalised results. However, the calculations may take a long time. The Conjugate Gradient and Quasi-Newton only use the first derivative of the function. Therefore, these methods are regularly preferred in applications when only the first derivative is known or when higher results are expensive to calculate [72].

3.10 Ensemble Classifier

Many researchers have shown that combining a set of different classifiers with different misclassified instances will generate better classification performance compared to a single classifier, which would develop the ensemble system having peak performance [73]. The concept is that if individual classifiers make errors on different instances, combining these different classifiers can reduce the overall error to improve the performance of the ensemble system [74].

The success of an ensemble classifier depends on having a diversity between individual classifiers concerning misclassified instances. There are four different ways to achieve better accuracy performance in the ensemble classifier [74]:

1. Use different training instances to train individual classifiers
2. Use other training parameters when tuning the classifiers
3. Make use of extra features to train the classifier
4. Combine the selected classifiers

The first ensemble classification method introduces different resampling techniques, including ageing and boosting [75], [76]. The second approach that may be considered is implementing different parameter values such as weights, nodes, or layers to train the classifier. The third approach deals with operating with various features to train the classifier, whilst also a combination of different classifications may also be used.

Lee et al. [77] proposed a breed of context prediction mechanism using the Markov Blanket obtained from the General Bayesian Network (GBN) as the primary vehicle. An ensemble of robust prediction classifiers was suggested to improve the prediction accuracy. Entirely different classifiers were used to construct ensemble systems for location prediction experiments. Lee et al. selected three types of individual classifiers – decision trees, Bayesian classifiers, and SVM – and integrated them using two different combination strategies – voting and stacking.

There are three significant steps to produce an ensemble learning technique, regardless of the procedure [78].

- **Ensemble Generation**: this phase is used to create a few samples, each of which constructs a classifier utilising a single learning model.

- **Ensemble Pruning:** eliminates some of the classifiers that have been created in the beginning (first step). The aim is to decrease the total size of the tree without affecting the accuracy or performance.
- **Ensemble Integration:** This method uses a voting or averaging strategy to combine the models to predict any new cases.

3.11 Evaluation Metrics Techniques

Performance evaluation metrics are vital in machine learning algorithms when estimating a classifier or ensemble classifier [79]. Many techniques are used in this study which will be discussed further. Some researchers have suggested that accuracy and false positive rate are great at estimating the error rate classification. However, other researchers such as Davis et al. [80] and Kotsiantis et al. [81] proposed that accuracy and false-positive are not enough to generate accurate results. ROC, AUC, precision, recall, accuracy should also be implemented as an evaluation metric.

3.11.1 Confusion Matrix

The evaluation techniques are conducted using a confusion matrix. Figure 3.11 represents the confusion matrix. There are four donates that are in the contingency table. True Negative (TN) and True Positive (TP) are the negative and accurate classification of positive instances, respectively. False Negatives (FN) illustrate the positive values incorrectly classified as negative. In contrast, False Positives (FP) show negative values poorly associated with positive ones [82]. Some equations can be used to find the performance evaluation measurements:

Sensitivity is the percentage of positive instances out of the ***actual positive*** results. Therefore, the denominator (TP + FN) is the total of the positive cases presented in the dataset, as shown in Equation 3.4 [83].

$$Sensitivity = \frac{TP}{TP+FN} \qquad (3.4)$$

Specificity is the percentage of negative instances out of the ***total negative*** cases. Therefore, the denominator (TN + FP) is the number of negative instances present in the dataset, as shown in Equation 3.5 [83].

$$Specificity = \frac{TN}{TN+FP} \qquad (3.5)$$

Precision is the percentage of positive instances out of ***predicted positive*** cases (true or false positive predictions). In another way, the rate of sending non-spam emails to the junk mail. Equation 3.6 shows the calculation of precision [83].

$$Precision = \frac{TP}{TP+FP} \qquad (3.6)$$

Recall goes in another direction; it measures the ratio of True Negatives against the total of predicted negatives (whether true or false ones). It can be calculated using Equation 3.7 [84]:

$$Recall = \frac{TN}{TN+FN} \qquad (3.7)$$

F-Measure is the average of precision and recall. Therefore, the higher the F1 score received, the better the accuracy. A model is found to do well in the F1 score if the positive predicted are positives (precision) and does not miss out on positives and indicates them negatively (recall), as shown in Equation 3.8 [83].

$$F1\ score = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} = \frac{2 \times precision \times recall}{precision+recall} \qquad (3.8)$$

Accuracy is the most usually used metric to judge a model; it is the rate of the total correct predictions against the total ever results. However, accuracy is not a clear indicator of the model's performance, specifically when the classes are imbalanced, according to Equation 3.9 [83].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \qquad (3.9)$$

The ROC curve is an evaluation metric for binary classification problems. The probability curve plots the TPR against FPR at various threshold values and separates the signal from the noise. The area under ROC Curve is a performance metric for measuring the ability of a binary classifier to discriminate between positive and negative classes [83]. Figure 3.6 illustrates the confusion matrix.

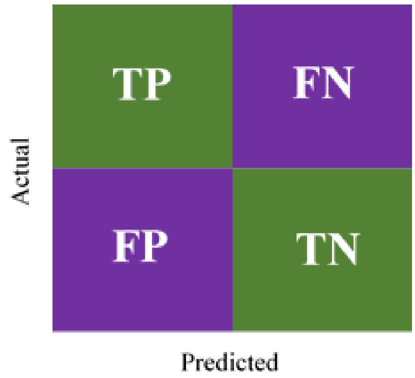


Figure 3.6 Confusion matrix [83]

3.12 Chapter Summary

This chapter provides details about machine learning algorithms used in the study. The computational and mathematical techniques are discussed in detail for each model to provide a clear idea about the procedure of the algorithm itself. This chapter also presented the combination of models to deliver better outcomes, indicated in the literature review chapter. In addition, the performance evaluation metrics are being discussed in brief in this chapter, the evaluation techniques are conducted using confusion matrix.

Based on the conducted literature review, this research used Neural Network classifiers trained using Levenberg-Marquardt Neural Network Algorithm (LEVNN), Backpropagation Network Classifier (BPXNC), and Random Forest Classifier (RFC). The next chapter will discuss the proposed methodology and experimental setup for the modelling environment and dataset pre-processing.

Chapter 4 Proposed Methodology

4.1 Introduction

This chapter discusses the proposed framework and the experimental setup design to solve some of the identified issues in the literature review relating to student levelling. Several studies applied machine learning models to aid teaching and learning in lower and higher education. Those studies are discussed in chapter two. In the literature review of student levelling, researchers have addressed the difficulty of classification in student levelling that has already been suggested or implemented. The primary purpose of this chapter is to build on the contributions hence develop the required outcomes by generating a novel framework that has not been applied yet while also viewing machine learning in student levelling from a different perspective.

4.2 The proposed framework

The proposed framework and experimental set-up used applies several individuals and combined algorithms. The study also investigates details on the methods of pre-processed data, feature selection, classification techniques, combined classifiers, and evaluation approaches to check the performance and accuracy of the tested models.

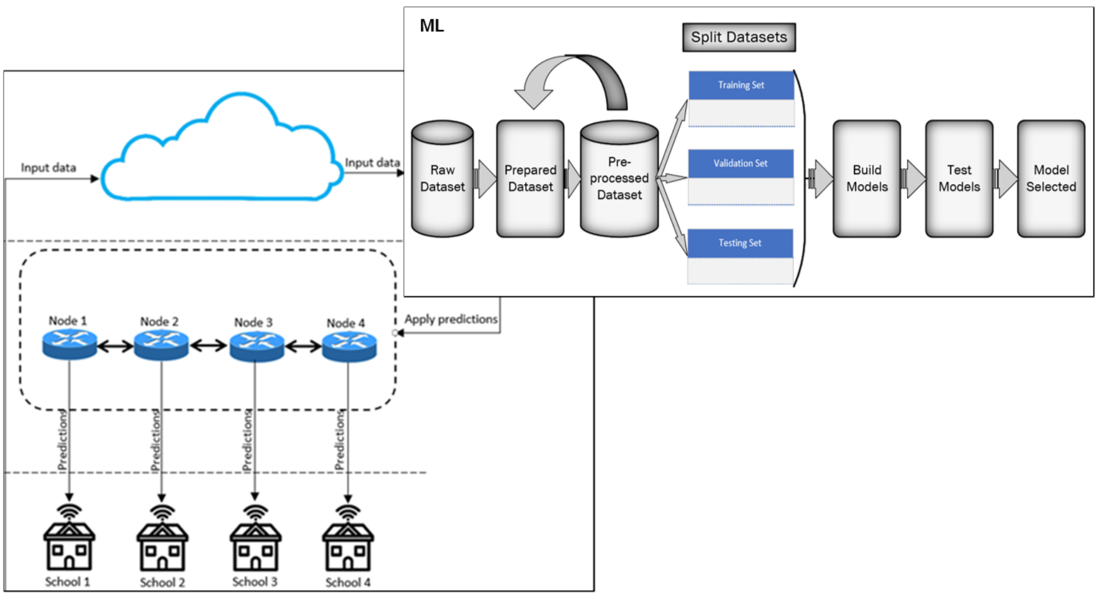


Figure 4.1 Proposed Framework

The methodology discusses the student dataset collected from schools, including the student exam marks vital for this study. Figure 4.1 describes the proposed framework for the research technical work.

Once X data is inputted to the browsers to visualise a real-time scenario, the routing and controllers will pass the student's information to the modals and store it in the database. When the student's data is subjected to levelling, the request is sent to the server again. The data gets downloaded back to the school nodes and the brain.js, which is now subjected to the pre-defined algorithm of decision making, which will push out the students' results and await the user's confirmation for levelling. Once the proof is received, depending on the level of access to the school, the data gets sent back to the routes, controllers, modals and finally stored back in the database as different datasets. The dataset stored in the database is connected to the machine learning algorithm in the TIBCO tools (subjected to program the ML). Once the ML is programmed with the preferred algorithms, the datasets will be processed with different parameters and the algorithm will be trained with the training dataset.

Despite the focus of this study being based in the UAE, the dataset used was collected from international schools allocated in Abu Dhabi, UAE. Therefore, it does not cover all curricula around the world. Yet, the method can still be used by schools, teachers, and the Ministry of Education to provide accurate student levelling and provide predictions for the future. Predominantly, most schools in the UAE assess their students by written exams and regular evaluations by teachers. Yet, each curriculum and exam board follow a different approach to student levelling [85].

Several practical machine learning approaches such as RFC, ANN models, and combined classifiers have been used in this research. The aim is to obtain outcomes with the lowest tolerable error rate. This research applied a stacked generalisation method to improve accuracy by combining several algorithms and the learning ensemble classifier of a specified dataset [86]. This study evaluated the results obtained from the ensemble and single classifiers and revealed which model provided better outcomes. In addition, various quality measures are used to assess and benchmark multiple machine learning algorithms [87]. The following sections will discuss the proposed methodology and machine learning implementation techniques for student levelling.

4.3 Ethical Approval

There are several ethical challenges and considerations regarding handling actual student data. Ethical approval needs to be obtained from Abu Dhabi Department of Education and Knowledge (ADEK) to research schools [85]. The permission to access student data was planned in the earlier research stage and was granted by LJMU's Research Ethics Committee. In addition, gaining a letter of authorisation from ADEK was crucial for the research. All researchers of schools in Abu Dhabi must have this permission beforehand. All the required ethical approval supporting documents were submitted to the UREC research ethics team, including ADEK letter of authorisation, LJMU research ethics training, research protocol, participant information sheet, gatekeeper information sheet, and the consent forms (UREC reference: 19/CMS/001) [88].

Participants were informed clearly in the participant information sheet about the purpose of the research and the aim before providing the researcher with the required information. It was clearly explained to participants in writing that all provided information would be confidential and for the use of the research [89]. Their completion of the written consent form granted the researcher approval to research with that participant [89].

There were risks anticipated before commencing data collection [90]. One of the risks that were taken into consideration is that a written consent form needs to be provided by participants to permit the researcher to conduct the research, which would be tricky [91]. Another risk was through the document analysis, as some documentation was challenging to access or protect from the public. Information on children and their backgrounds can be sensitive; this was an anticipated difficulty [91]. Choosing the key persons to allow the data to be varied depends on having school managers and decision-makers participants. As this is quantitative and qualitative research, purposive sampling is intended to be used in this research. Sampling is conducted specifically to fulfil this purpose: collecting data from two central education systems in the United Arab Emirates: British and American. The decision has been made to whom to interview allows an extensive set of data to be collected.

Privacy concerns surround the web-based system's handling of students' data. It is vital to ensure that the data collected for the research is anonymised. According to the Data Protection Act, participants are not directly attributed to data and information disseminated [99] [100].

4.4 The Proposed Methodology

Machine learning models have been utilised to classify student academic levels and e-learning. Using the proposed ML model, schools can predict students' current levels and predict their educational path in the future [92]. The main benefit of this model is to use recent technological development in ML algorithms to assist schools and teachers in developing their students more effectively with the help of records [93]. The aim is to propose the prediction of the student level using classification methods based on the dataset collected from schools in the UAE.

The proposed implementation comprises several steps; data collection, data pre-processing, and then building the model based on the training data, evaluating the model based on the testing sets, and selecting the relevant model. The data collection process starts when the student is first registered. The collected data must be cleaned (removing unwanted data and filling in missing data). After that, various ML models are selected to evaluate the data sets [94]. The holdout method is used to split the dataset into training, validation, and testing, as shown in Figure 4.2 [95].

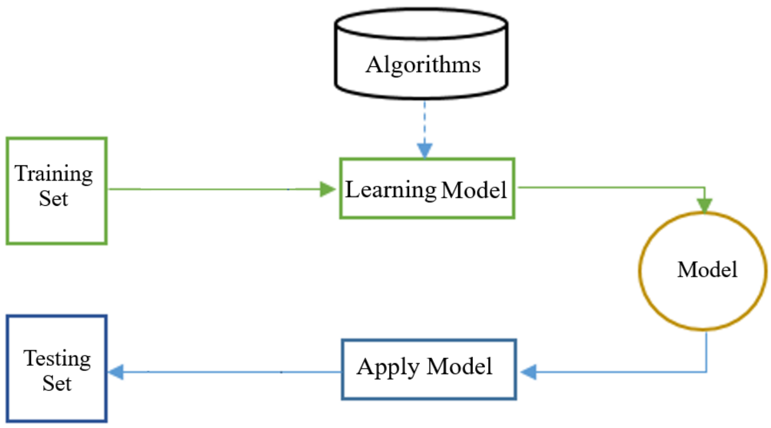


Figure 4.2 The methodology process

Learning-based classifiers' critical feature is their ability to adjust the internal structure depending on the input and the respective target value (desired output) [96]. Overall, student admission, levelling, and differentiation are not consistent throughout schools as some use their levelling criteria while others use external agencies. Therefore, the suggested model is to use Machine Learning to provide levelling and differentiation information of students who relocate from a particular education curriculum to another whilst having the ability to store and access student data from anywhere throughout their academic journey. Table 4.1 indicates the main parameters and models used in this simulation experiment study [97], [98].

Table 4.1 List of parameters used for the proposed model

No	Type	Number	Description
1	Data instances	1550	Data collected for two academic years
2	Class Variables (Output data)	3 Classes 1 Feature	Class 1 is 85% ≤ 100% Class 2 is 75% ≤ 84.99% Class 3 is < 75%
3	Features (Input data)	30 Features	Student Name, Student ID, Gender, Date of Birth, Proposed Year/Grade, Year of Admission, Previous Curriculum, Current Curriculum, Previous Year/Grade, Math Entry Exam Mark, Science Entry Exam Mark, English Entry Exam Mark, Maths Marks 19-1, Science Marks 19-1, English Marks 19-1, Maths Marks 19-2, Science Marks 19-2, English Marks 19-2, Maths Marks 19-3, Science Marks 19-3, English Marks 19-3, Maths Marks 20-1, Science Marks 20-1, English Marks 20-1, Maths Marks 20-2, Science Marks 20-2, English Marks 20-2, Maths Marks 20-3, Science Marks 20-3, English Marks 20-3
4	Evaluation Metrics of classification models	6	Sensitivity, Specificity, Precision, F1 Score, Accuracy, and Youden's J statistic (J Score) values.
5	Visualization Techniques	2	ROC curve and the Area Under the Curve (AUC)
6	Machine Learning Algorithms	4	LEVNN, RFC, and BPXNC
7	Ensemble Classifiers	4	LEVNN1 COM, LEVNN2 COM, and LEVNN and RF COM1, and LEVNN and RF COM2

The primary purpose of this research is to use the recent advances in machine learning models to assist schools and teachers in assessing their students' levels, predicting their levels, and storing their levelling data to predict feature outcomes in advance. It can potentially be more effective and efficient for schools to level students and aid in student transfers between schools whilst maintaining or improving the level of the student. The remainder of the chapter will discuss these processes within the proposed framework procedure.

4.5 Raw Data Preparation Process

The dataset collected from different schools must be complete and coherent for the model to generate optimal results. There are two steps to be taken as stated below to prepare this dataset:

4.5.1 Description of the Raw Data

The dataset used for this study was collected from two schools, each with a diverse curriculum which is British and American curricula, for two academic years. Table 4.2 and Figure 4.3 show the number of the records in each of the three classes used in the research, which been collected from the schools. For the model to generate optimal and reliable results, much data is required, provided by the school gatekeepers. The marks for each subject (Maths, Science, and English) were taken as an average for each term and combined both academic years to aid the classification process. The class division was performed to provide adequate class representation over the data sample while preserving a decent margin between the set subject mark values. Since the data sample consists of 1,550 records, having more than three classes decreases the total records in each category and reduces the correlation between the collected samples [99].

Table 4.2 Number of records in each of the three classes

No	Classes	Class Description	Total Record in Each Class
1	Class 1	85% ≤ 100%	480
2	Class 2	75% ≤ 84.99%	1016
3	Class 3	< 75%	51

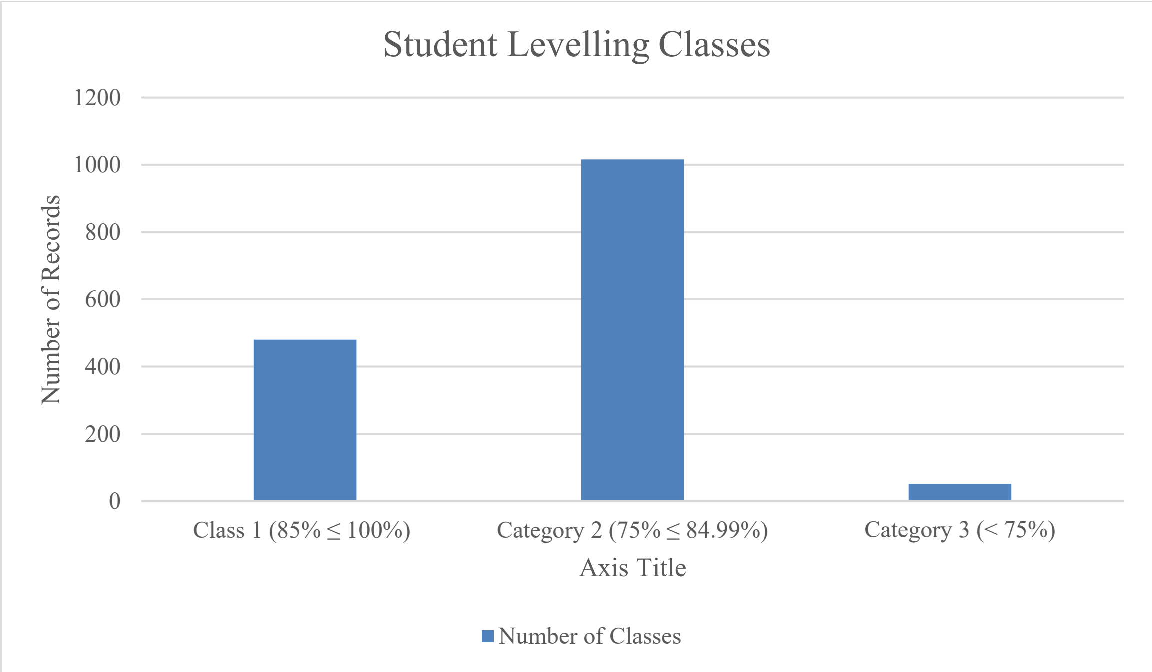


Figure 4.3 Number of records for the three classes

4.5.2 Data Attributes

The dataset presented comprises students' records for two academic years that include Math, English, and Science courses for three terms and the entry exams results of these courses. Previous researchers influenced the selection of subject areas and some terms in this research in a similar subject matter. The dataset comprises novel aspects, specifically student grading in diverse educational systems within a country. Each sample consists of 34 attributes, as shown in Table 4.3.

Table 4.3 Attributes of student levelling dataset

#	Attribute Name	Value	Description
1	Gender	Male/Female represented by 0 and 1	Gender of the student
2	Student Age (As of 2017/18)	6,7,8, 9, etc.	Age of the student calculated from 2017 / 18 academic year
3	Age Group	-1, 0, 1	-1 means one year below the legal age; 0 is on the legal age, and one is over legal age by one year.
4	Year of Admission	Old BBS Student, Old GEMS Student, New Admission 18/19	The data collected is for two or more academic years; 2017/18 and before academic years + 2018/19 academic year.
5	Current Year (17/18)	FS1, FS2, Y1-12 / Grade 1-11	The year or grade group assigned to the student by the school
6	Proposed Year/Grade (18/19)	FS1, FS2, Y1-13 / Grade 1-12	The year or grade group assigned to the student by the school
7	Year of Admission	Old BBS Student, Old GEMS Student, New Admission 18/19	The data collected is for two or more academic years; 2017/18 and before academic years + 2018/19 academic year.
8	Previous School (17/18)	Many schools in UAE	Previous schools that the student was in before this study
9	Previous Curriculum	UK / US / MOE / Canadian / Indian / Australian / CBSE / German	The curriculum the student transferred from.
10	Current School	GEMS, BBS	Name of the school that the data has been collected
11	Current Curriculum	0, 1	The curriculum that the student transferred to, 0 = US, 1 = UK
12	Math-exam	Mark out of 100	Exam marks for school entry exam in math
13	Science-exam	Mark out of 100	Exam marks for school entry exam science
14	English-exam	Mark out of 100	Exam marks for school entry exam English
15	Math19-1	Percentage out of 100%	Term 1 student Maths Exam marks during the academic year 2018/19
16	Science19-1	Percentage out of 100%	Term 1 student science Exam marks during the academic year 2018/19
17	English19-1	Percentage out of 100%	Term 1 student English Exam marks during the academic year 2018/19
18	Math19-2	Percentage out of 100%	Term 2 student Maths Exam marks during the academic year 2018/19
19	Science19-2	Percentage out of 100%	Term 2 student science Exam marks during the academic year 2018/19
20	English19-2	Percentage out of 100%	Term 2 student English Exam marks during the academic year 2018/19
21	Math19-3	Percentage out of 100%	Term 3 student Maths Exam marks during the academic year 2018/19
22	Science19-3	Percentage out of 100%	Term 3 student science Exam marks during the academic year 2018/19
23	English19-3	Percentage out of 100%	Term 3 student English Exam marks during the academic year 2018/19
24	Math20-1	Percentage out of 100%	Term 1 student Maths Exam marks during the academic year 2019/20
25	Science20-1	Percentage out of 100%	Term 1 student science Exam marks during the academic year 2019/20
26	English20-1	Percentage out of 100%	Term 1 student English Exam marks during the academic year 2019/20
27	Math20-2	Percentage out of 100%	Term 2 student Maths Exam marks during the academic year 2019/20
28	Science20-2	Percentage out of 100%	Term 2 student science Exam marks during the academic year 2019/20
29	English20-2	Percentage out of 100%	Term 2 student English Exam marks during the academic year 2019/20
30	Math20-3	Percentage out of 100%	Term 3 student Maths Exam marks during the academic year 2019/20
31	Science20-3	Percentage out of 100%	Term 3 student science Exam marks during the academic year 2019/20
32	English20-3	Percentage out of 100%	Term 3 student English Exam marks during the academic year 2019/20
33	Average 19/20	Percentage out of 100%	The average of 2019/2020 academic year extracted from the marks of the three subjects among the three terms
34	Class	1, 2, and 3	3 classes extracted from the Average value; class 1 from 85% to 100%; class 2 from 75% to 84.99%; and class 3 is below 75%

4.6 Exploratory Analysis of Dataset

Exploratory analysis is an essential step that must be considered when dealing with data and its learning ability [100]. The information from the data exploration can influence the design of the modelling phase as most of the learning ability. The investigation of the utilised data in these simulations is achieved by having statistical data alongside visual charts, which includes t-distributed Stochastic Neighbourhood Embedding (tSNE) and Stochastic Proximity Embedding (SPE) [101], [102].

4.6.1 T-distributed Stochastic Neighbourhood Embedding (T-SNE)

Data representation is examined to see any patterns within its structure [103]. Furthermore, the exploratory step is intended to expose any outliers and other questionable aberrations in the data, if any exist, so that the results of the subsequent analysis are not tainted by faulty input [104]. Exploratory research is crucial in the machine learning process because it allows the human adviser to understand the data and its potential to learn [105]. Since a critical component of learnability is known to be a function of the correspondence between the learning algorithm and the data, the outcomes of data exploration can be utilized to influence the modelling phase [105].

Figure 4.4.4 shows the student levelling dataset with 3 class labels. The plot illustrates the class dispersion problem with different colours, where points from the three classes of the student dataset are clustered. Ideally, the three classes are decomposed using a clustering technique; each cluster can determine a new class label for the testing set. The plot shows using T-distributed Stochastic Neighbourhood Embedding (tSNE) of the class distribution problem: groups with the same class points are spread across the variable values [103]. The purpose behind using t-SNE is to show dimensionality reduction that aids in visualising the student dataset with high dimensions.

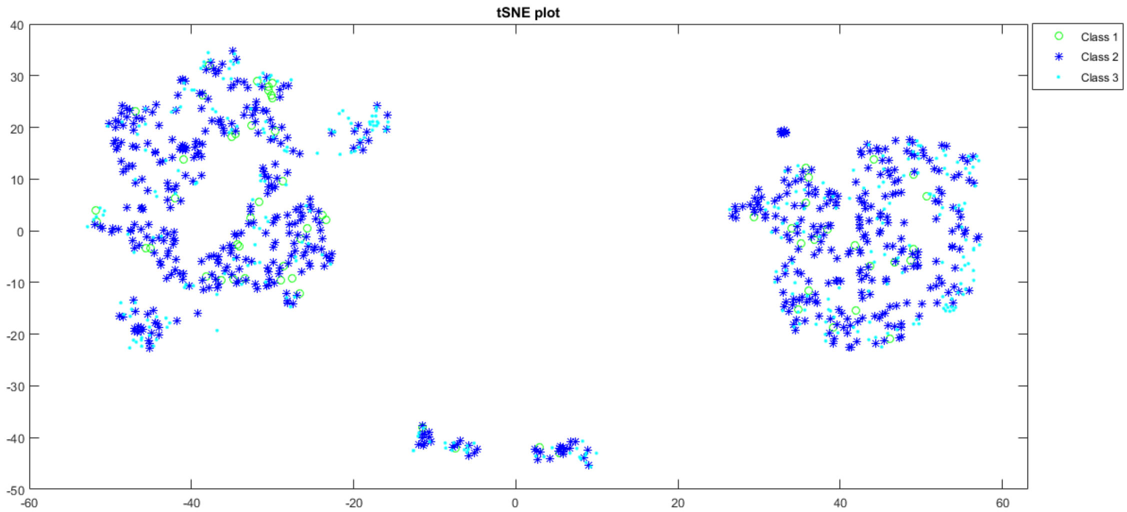


Figure 4.4 tSNE Plot for the Educational dataset

4.6.2 Stochastic proximity embedding zoom out (SPE)

Stochastic Proximity Embedding (SPE) is presented as shown in Figure 4.5, which is considered a novel self-organizing algorithm for constructing substantial underlying dimensions reduction from proximity data [102], [106]. SPE aims to generate low-dimensional embedding with the most significant similarities between related observations [107]. SPE creates an initial configuration, then selects pairs of objects randomly and modifies their coordinates in terms of their distances on the map according to their respective proximities [108].

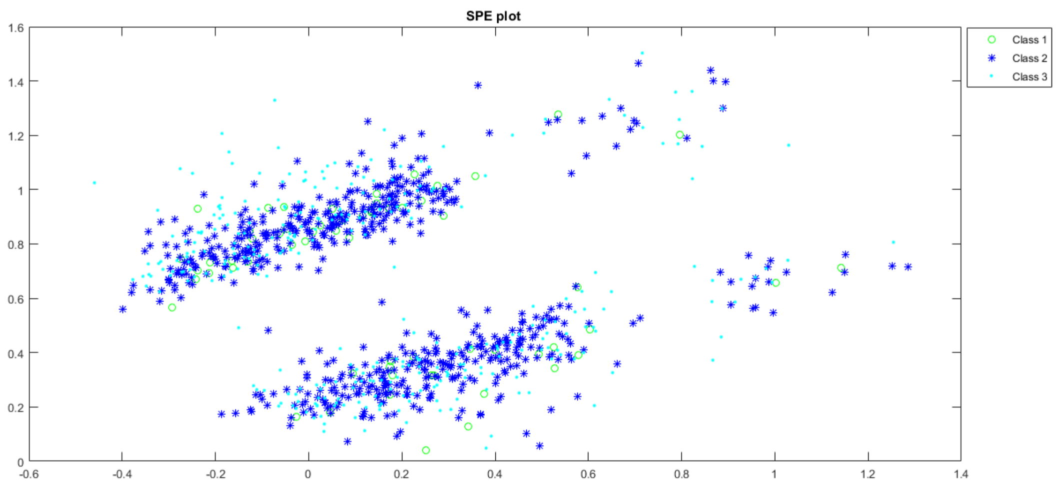


Figure 4.5 SPE Plot for the Educational dataset

4.7 Pre-Processing Techniques

Data processing is essential when using machine learning to classify or predict features in the dataset. This data processing method converts the raw dataset into a clean dataset that can be applied to the ML models. Without data cleaning, the models cannot process the data effectively; inaccurate, contaminated, inconsistent, and incomplete data analysis will lead to unreliable results [109].

The original dataset was collected, and the first cleansing step involves filtering to equalise the instances according to their categories. The dataset then needs to be cleaned and transformed into a suitable form to receive accurate outcomes from having the model process efficiently. The data collection from multiple schools leads to various data formats and missing values. Therefore, reducing noise and adding or removing missing values will improve accuracy and performance.

4.7.1 Data Cleaning

Data cleaning is an essential part of handling data that helps check for correctness, meaningfulness, and the security of data to be used [110]. Therefore, the data was cleaned by unifying the outputs, filling in missing data, and standardising the values according to categories.

4.7.2 Outliers Detection

Data mining, also referred to as anomaly detection, identifies observations that do not fit into the pattern of the dataset or do not share similar items in the dataset [111]. Outliers in the dataset are split into two categories, multivariate and univariate. The multivariate method is discovered in n-features (n-dimensional) based on Mahala Nobis distance. The research deals with a dataset having a wide range of n-features. Therefore, it is essential to distinguish the outliers.

On the other hand, univariates were discovered in a single feature space [112]. Outliers on students levelling datasets can occur when data from different schools with different levelling criteria and errors during data entry. Consequently, it is vital to use effective techniques to predict the outlier factors and replace them with good figures, as shown in Figure 4.6 of an example of the English Preparation Exam, which had 12 detected outliers and been replaced automatically. Other attributes outliers with related codes are added to the appendices.

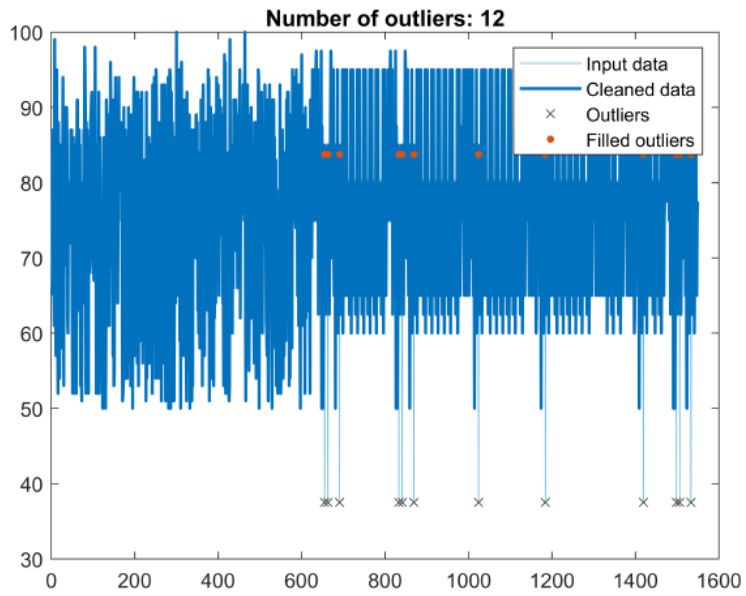


Figure 4.6 Outliers' detection example (English Preparation Exam)

4.7.3 Missing Values

Missing values and missing features in the student levelling dataset are common issues faced with data preparation [113]. Having features with missing values can be an issue because the feature might become less effective and biased [114]. Missing essential values in the student dataset could happen due to unavailability of data, gatekeeper confusion, and some schools may not be willing to provide this data. Missing data can also occur if the gatekeeper does not provide the required information on time, and therefore the lost data can lead to biased results [115].

Several algorithms can handle missing data by ignoring them. However, most algorithms need to have the data cleaned and completed before implementing them into the models. As a result, missing values can concern classification and regression models as the algorithms cannot function. Therefore, missing values must be identified and resolved to have a functional classification [112].

4.7.4 Missing Data Mechanism

There are viral factors that must be addressed when faced with missing data. To discover the amount of missing data being dealt with, the exploratory analysis must be implemented to distinguish the impact of the missing data. It should be noted that missing data ranging between 2% to 3% will not significantly impact the model. Otherwise, this issue must be solved [116].

4.7.5 Data Integration and Normalisation

Data integration is a method that draws data from different sources and combines them into a single database. The data integration method processes data into a single dataset compatible with machine learning models [117]. Identifying and resolving data errors is essential, so the data integration process fixes those errors with different values, attributes, and formats [118].

Normalisation is the optimal option used for the transformation of the data structure. There are several data normalisation approaches, including statistical and arithmetic rules. Most normalisation is done by converting values of quantitative features to two values such as 0, 1 or -1, 1 [117]. This study applied the Normalisation process, in which the dataset was uploaded, and the values were prepared by converting them into a numeric format, as shown in Figure 4.7.

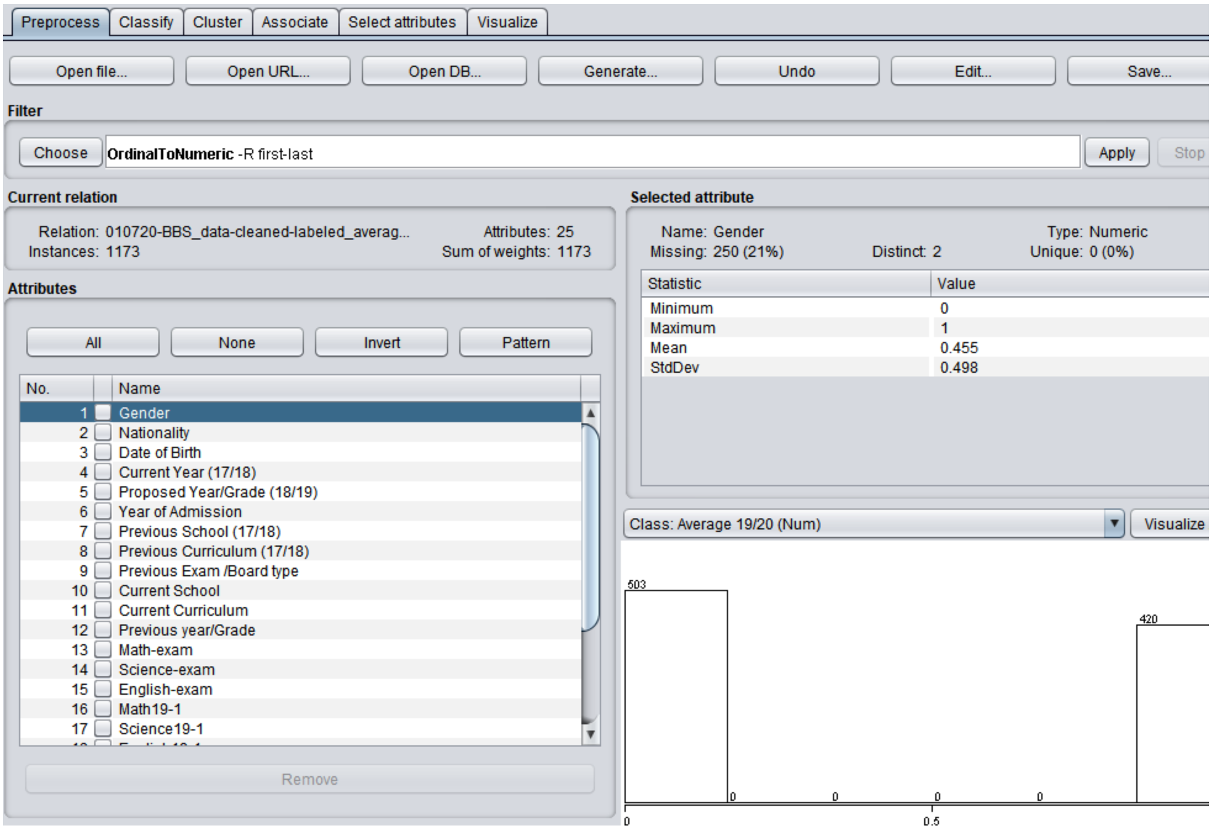


Figure 4.7 Converting dataset values to numeric

Secondly, normalising the dataset as shown in Figure 4.8

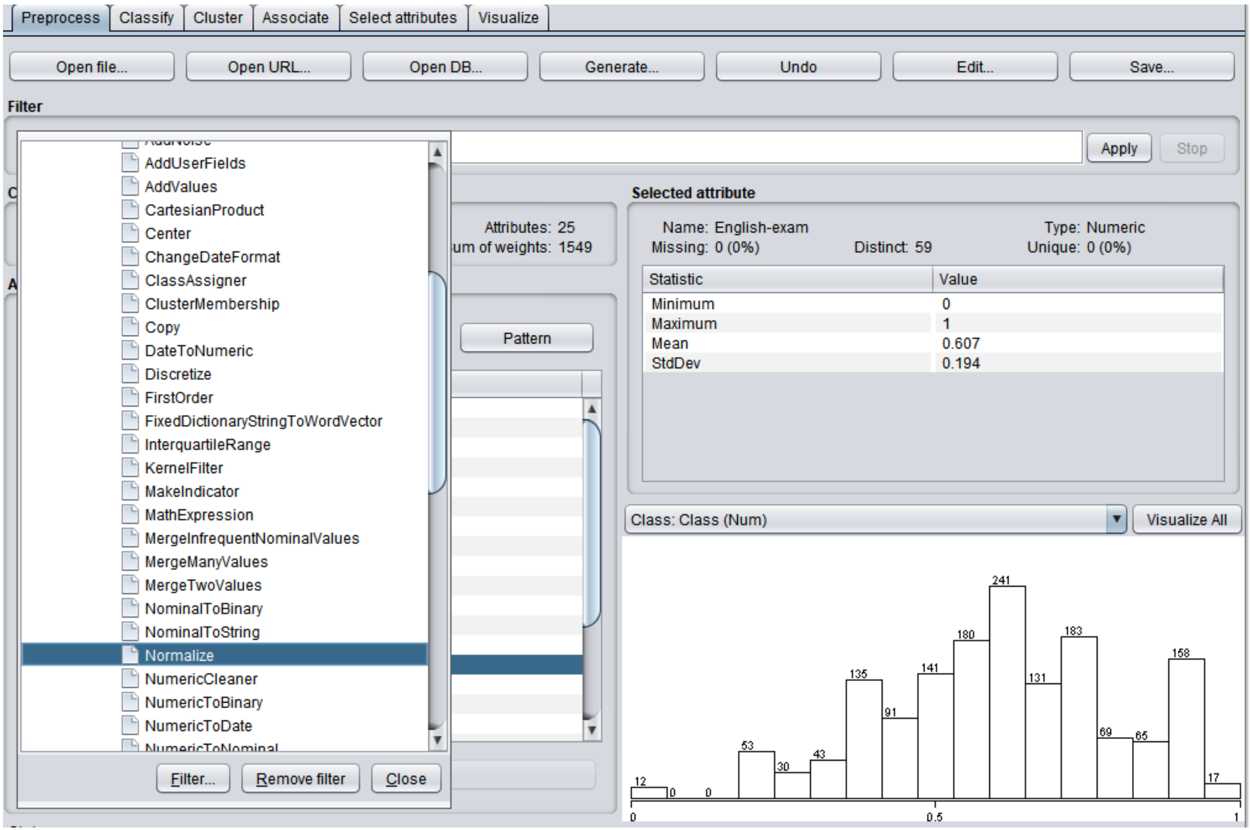


Figure 4.8 Normalising the dataset

Finally, standardizing the dataset as shown in Figure 4.9

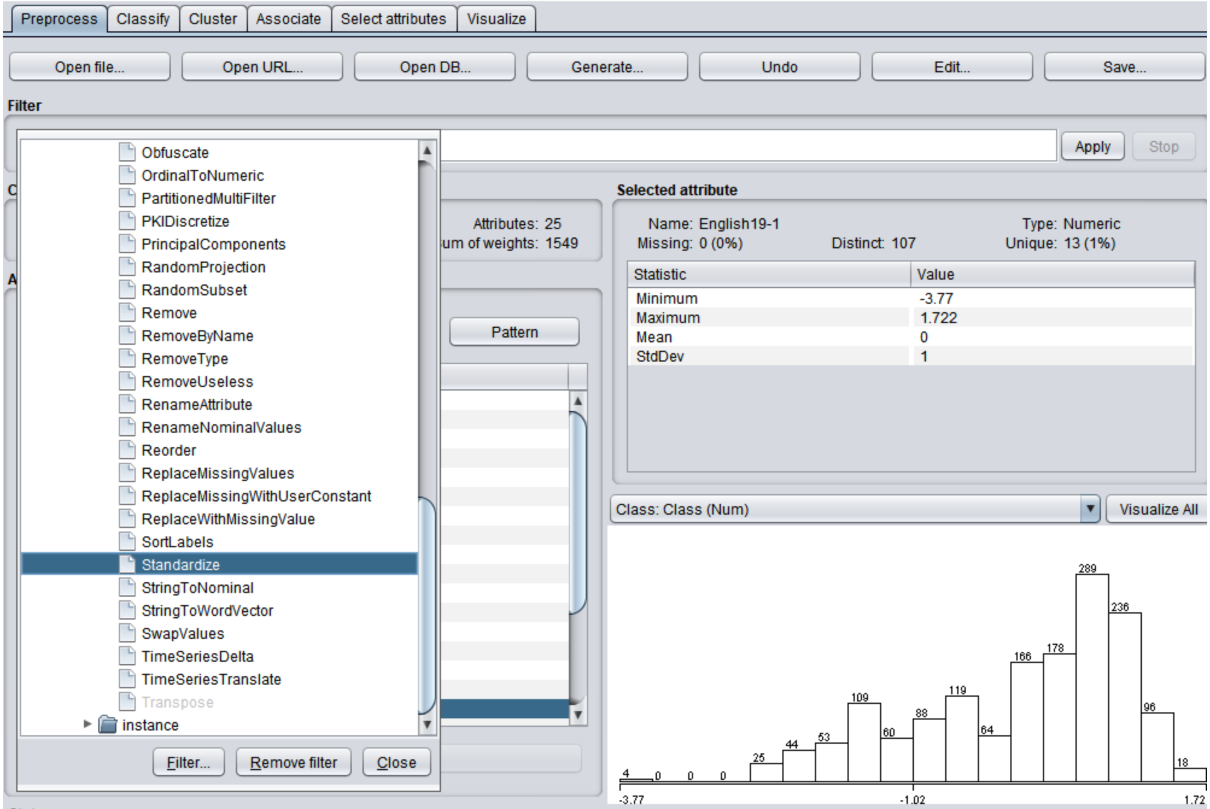


Figure 4.9 Standardizing the Dataset

4.7.6 Feature Selection

Feature selection is one of the essential pre-processing steps in data mining [119]. The feature selection is to select a subset by disregarding irrelevant features and unwanted information from the student levelling dataset. It is an effective dimensionality reduction technique to remove noise features. In general, the basic idea of a feature selection algorithm is to search through all possible combinations of attributes in the data to find which subset of features works best for prediction. Thus, the attribute vectors can be reduced by which the most meaningful ones are kept, and the irrelevant or redundant ones are removed and deleted [119].

Feature Subset selection has two approaches: Filter and Rapper. The filter approach applies data with an examining property, in general, to calculate the goodness of the feature subset except for a learning algorithm that evaluates the quality of the feature subsets [120]. Using this technique in the research, unnecessary features were reduced. Isolating irrelevant features helped improve the performance of the models and the results generated from the dataset. Overfitting can harm the performance of the models; therefore, this technique can reduce those risks [121]. Feature selection decreases the search space determined throughout all the features, and consequently, the models can process the data faster with less memory consumption [122] [123]. The irrelevant and redundant features can confuse the learning when dealing with a few training examples, leading to overfitting and high dimensionality [124]. The high dimensionality of the extracted features will be reduced using feature selection methods [125]. This method is achieved by identifying spaces with lower dimensions than the actual data.

The process of feature selection is divided into two forms [126]:

(i) Feature transformations – dealing with lower dimensional space such as independent component analysis (ICA) and principal component analysis (PCA).
(ii) Select some features for a given pattern based on the mean or standard deviation of the feature values.

Feature selection can reduce both the data and the computational complexity [127]. Various feature selection methods are available such as Information Gain (IG), Symmetric Uncertainty (SU) and Correlation-based feature selector (Cfs). Figure 4.10 illustrates how the Cfs works.

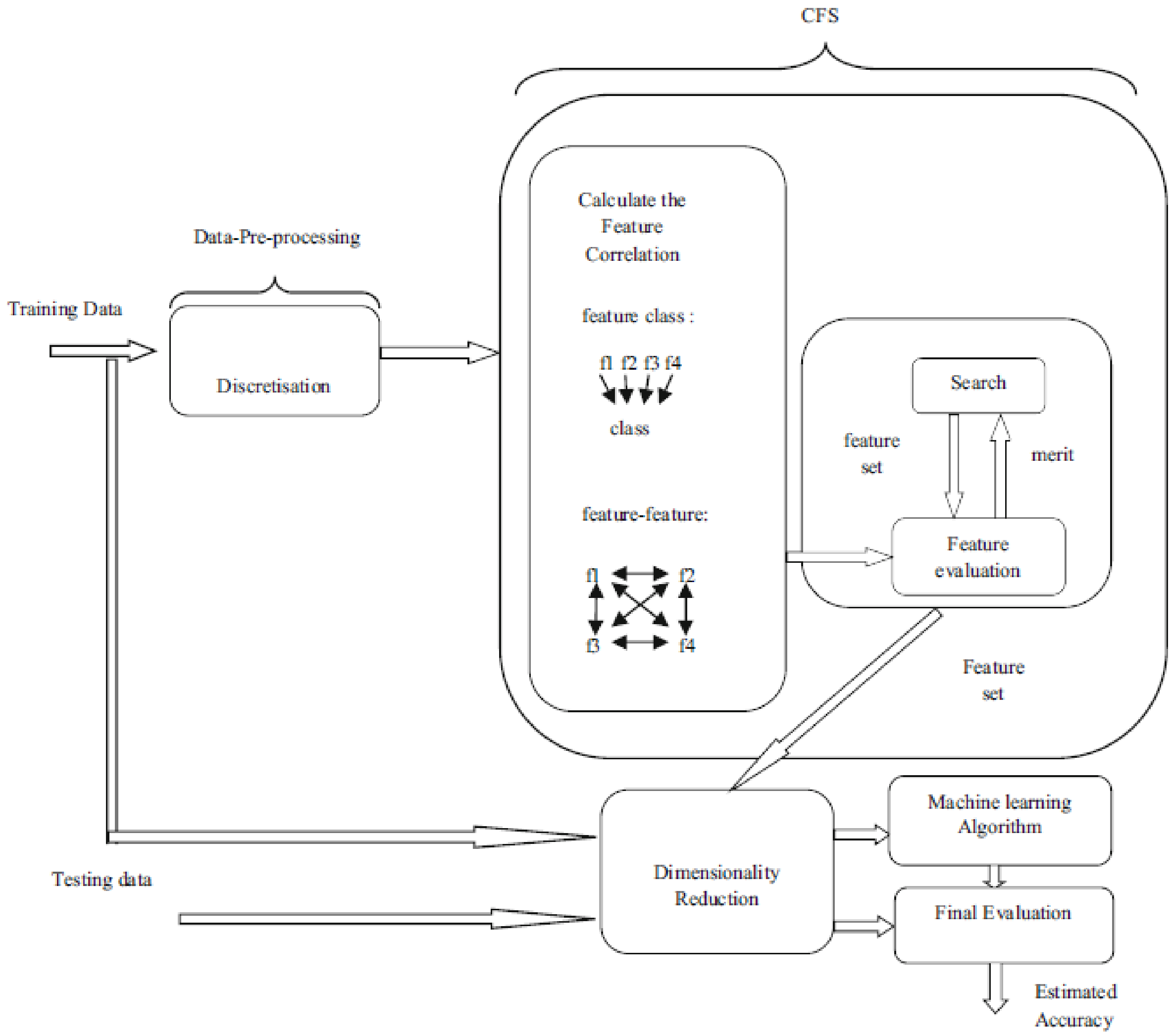


Figure 4.10 Correlation-based feature selector (CFS) [120]

In this research, names and student IDs should be removed from the dataset to avoid disclosing students' personal information as per Ethical Approval. Then, the date of birth needs to be converted to age for categorisation. A new attribute of age group having three categories: -1, 0, and 1, was created according to the student's age and legal age of the class the student was assigned to. This process will allow the Age Group feature to be used instead of the Date of Birth, so Date of Birth and Age features will be excluded from this research.

The Previous School attribute has been removed because it has the same effect as the Previous Curriculum attribute. Nevertheless, the Previous Curriculum attribute will be replaced by three columns as below:

1. PrevCurrUS: 1 for the US curriculum, and all other curricula are represented by 0.
2. PrevCurrUK: 1 for the UK curriculum, and all other curricula are represented by 0.

3. PrevCurrOther: 1 for all curricula except the US and UK (e.g., Asia, UAE, Pakistan, Iraq etc.), and the US and UK curricula would be 0.

Hint: There should be only one value of “1” among the three features in all cases, and the remaining attributes must remain zeroes. In other words, if a student has 0 0 1 respectively, they had neither British nor American previous curriculum. While having 1 0 0 will indicate that they had an American curriculum. And so, if they have 0 1 0, they were in a British curriculum. Table 4.4 explains the values of the new three features.

Table 4.4 Explanation of new extracted features and their values

Value	PrevCurrUS	PrevCurrUK	PrevCurrOther
1	US	UK	Others
0	Others	Others	US or UK

In addition, the Average attribute has been created using all the students’ marks for the 2019/2020 academic year. Then the class label feature has been built called “Class”, which divides the average into three categories: 1, 2, and 3. Class one from 85% and 100%, class two from 75% to 84.99%, and class three is below 75%. Finally, the attributes used to create other features are excluded from the dataset, such as 2019/2020 academic year marks and their average. As a result, ML algorithms will be applied on twenty-four attributes, as shown in Table 4.5 Updated Features.

Table 4.5 Updated Features

#	Features Selected
1	Gender
2	Age Group
3	Current Class
4	Proposed Class
5	Year of Admission
6	Prev_US_curr
7	Prev_UK_curr
8	Prev_Other_curr
9	Current School
10	Current Curriculum
11	Previous System
12	Math-exam
13	Science-exam
14	English-exam
15	Math19-1
16	Science19-1
17	English19-1
18	Math19-2
19	Science19-2
20	English19-2
21	Math19-3
22	Science19-3
23	English19-3
24	Class Label

As shown in the below Figure 4.11, Cfs Subset Evaluator and the best-first search method have been used in this study to get the final feature set because they gave good results for the students' performance dataset. Thus, Cfs Subset Evaluator and best-first search are applied as the feature selection algorithm in this research. Table 4.6 represents the features selected using the Cfs Subset Evaluator and best-first search.

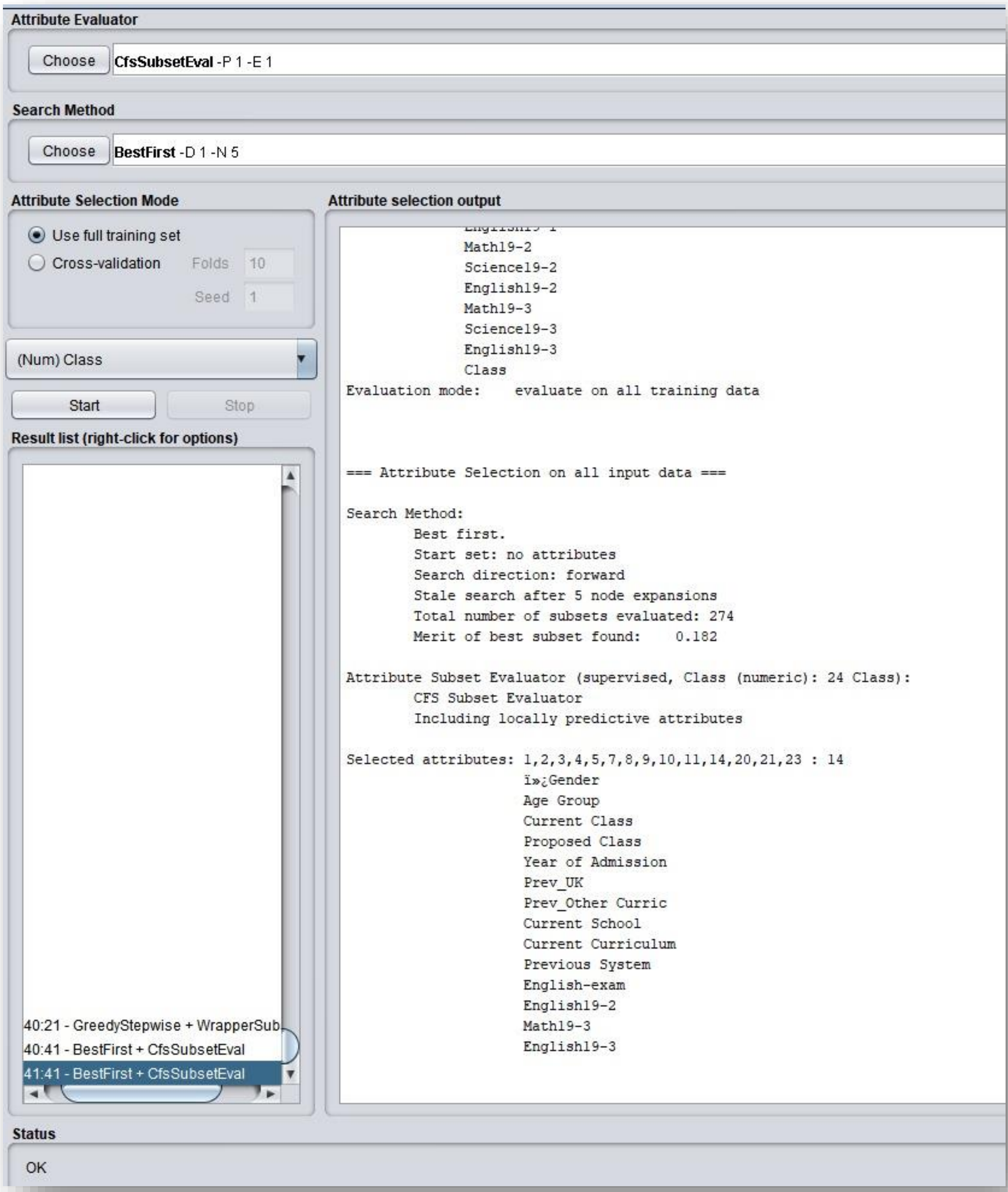


Figure 4.11 Feature Selection applied technique

Table 4.6 Features Selected Using the Cfs Subset Evaluator and Best First Search

#	Features Selected
1	Gender
2	Age Group
3	Current Class
4	Proposed Class
5	Year of Admission
6	Prev_UK
7	Prev_Other Curric
8	Current School
9	Current Curriculum
10	Previous System
11	Math-exam
12	English-exam
13	English19-2
14	Math19-3
15	English19-3

4.8 Experimental Setup

The experimental setup discusses the design of the test environment used in our experiment, the models used, and the configuration of each model [128]. The performance evaluation metrics utilised to measure the results of the machine learning algorithms are conducted for the student levelling dataset. The resulting dataset has 1550 samples, with a target variable predicting the level of the student in the future.

Two methods are used in this study of the implemented classifications, single and combined classifiers. Three single algorithms have been used in the research: LEVNN, BPXNC, and RFC classifiers. These models are considered robust non-linear classifiers and are suitable for comparing high accuracy and performance. Each model was tested repetitively, and the average of the results was calculated to obtain performance estimates for the respective models.

In addition, this research combined several machine learning models to get better performance and accuracy, which are NN Combination, LEVNN combination, and NN with RFC Combination (NN with RF Com 1, and NN with RF Com2). The testing results for the combined classifiers will be discussed in Chapter 5. Combining various classifiers will usually improve accuracy for the classification system [18].

4.8.1 Single Classifier Framework

As mentioned previously, this research applied three single machine learning algorithms: LEVNN, BPXNC, and RFC classifiers. These chosen classifiers are recognised as good at dealing with supervised datasets, so they have been selected for the research as they produced good results over the other classifiers. Each classifier will be explained thoroughly in the next chapter. Table 4.7 summarises the configuration of each model.

Table 4.7 Classification models description

Model	Description	Architecture	Training Algorithm	Parameters	Role
RFC	Random forest	14 inputs, 50 Trees, 3 outputs	Random feature bagging	The number of generated decision trees is 50. Size of feature: 1	Non-linear comparison model
LEVNN	Multilayer perception, Levenberg-Marquardt algorithm	Units 29-2-4 transit activations	Levenberg-Marquardt	Initialisation: Nguyen Windrow Adaptive learning rate settings: initial value: 0.001 coefficient for increasing LR: 10 coefficients for decreasing LR: 0.1 maximum learning rate: 1e10	Non-linear comparison model
BPXNC	The feed-forward neural network algorithm	Units: One context unit for each output unit	They are trained with mapping a set of input data to generate a computational modification for the whole weights.	Momentum coefficients between 0.01 and 1.0. Sigmoid function $f(x) = 1/(1 + e^{-x})$. Learning rate between 0.25 and 0.9. Performance: 0.0932	Non-linear comparison model

4.8.2 Combined Classifier

Machine learning uses a training set combined with building a classifier that provides a reliable classification [129]. This study used the multi-class classification problem where many classes are available in the dataset. This research combines multiple classifiers to improve the classification accuracy and performance compared to the single model. Different studies prove that doing so can generate better output [130]. The training set is delivered to each model. Each model generates an outcome using the performance metrics method. Voting was used to select the highest accuracy and performance classifier to discover the classifiers that create the highest performance.

The proposed research focuses on multi-class label classification where more than two classes are used in the dataset. Previous research has shown that machine learning algorithms with multi-class models generate better results with student datasets and have higher accuracy [131]. Therefore, this research proposes a multi-class label classification approach on student levelling dataset and discusses the performance of different methods that involve student performance data, as shown in Figure 4.12 [132].

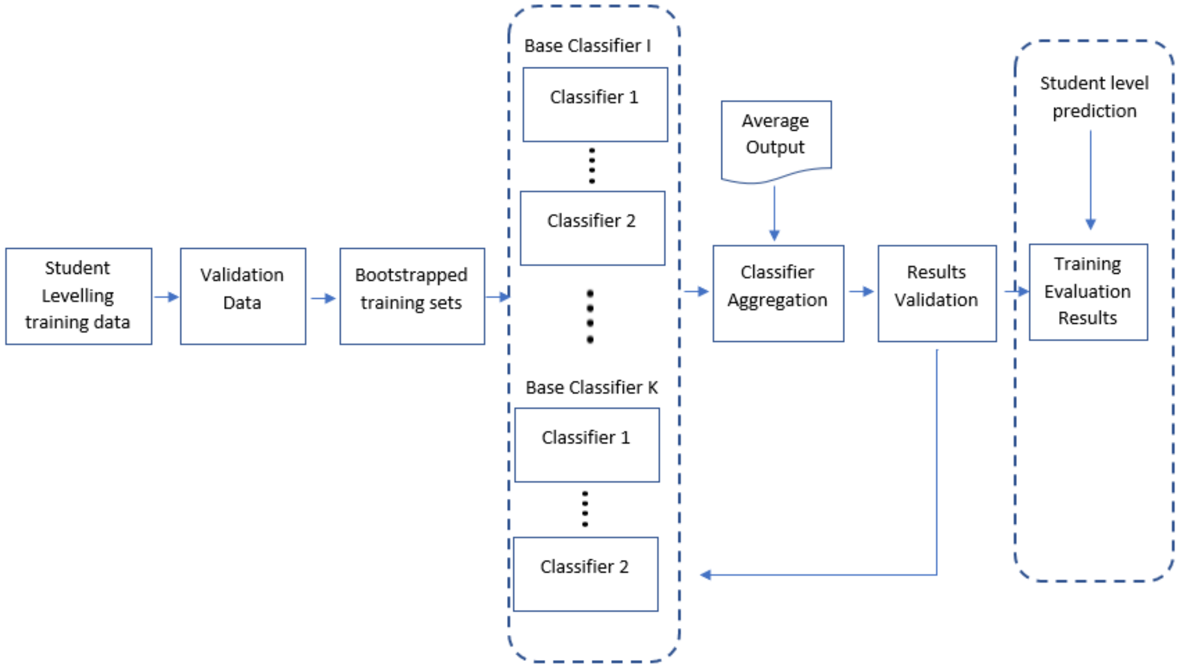


Figure 4.12 Training Phase Combined Classifier Workflow

The purpose of combining multiple algorithms is to create better results. The research used stacked and voting methods. The stacked process involves a set of models that lead to the same space combined. Each classifier receives training from the same training set, which receives 70% of the dataset, the validation gets 10%, while testing receives 20%.

4.9 Evaluation Techniques

Several methods are used for model evaluation. The primary purpose of model evaluation is to estimate the performance of the models (e.g., error rate and incorrect classifications) [133].

This research applied performance evaluation metrics by benchmarking the selected classifier outcome with the class attributes. The error rate, performance, and accuracy are calculated. The error rate for each model is computed by calculating the average number of misclassified instances divided by the number of features [134]. A threshold percentage for the classification accuracy shall be defined. If the accuracy is not achieved, then pre-processing methods, including feature selection, must be repeated using different ways until achieving better results [135]. Table 4.8 illustrates the most common approaches and their characteristics in machine learning algorithms.

Table 4.8 Evaluation techniques in machine learning

Evaluation method	Methodology	Description	Characteristics
K-fold Cross-validation technique [136]	Each classifier uses *n* – 1 group and holds one out of the fold for testing.	This method selects many folds (or divisions) to partition the data into each fold and is held out for testing. The process trains a model for each fold using all the data outside the fold. It tests each model performance using the data inside the fold and then calculates the average test error overall folds.	The outcome can be unbiased due to the *n* classifiers, the K-fold group is tested, and the *n* test outcomes are calculated.
Re-substitution	The total number of records in the dataset used for training and testing is equally.	All the available data was utilised for modelling to build a robust classifier.	The results generate biased estimation as the same data used for the training and testing process.
Holdout technique (Data Partition). (This method was implemented in the experiment).	Dataset divided between training sets and testing sets	The dataset is divided into training and testing sets. Usually, the training sets are received twice or more than the test size. In our thesis, the training sets receive used %70, the validation sets receive %10, while the testing phase obtains %20.	The model outcome estimation is unbiased in association with the error rates.
Jack-knife (Leave-out-one)	This approach typically has a similar function to k-fold cross-validation but *n* = N.	The classifier is very close to optimal because all samples get used for both training and testing.	The classifier result is unbiased but is considered slow concerning the computation-intensive task.

Zhou et al. [137] stated that the holdout method effectively selects a percentage of the available data with enough data for training, validation, and testing. Two-stage operations are required to build the learning scheme from the dataset. The error rate is calculated based on the created training method. The student levelling dataset is then evaluated using the testing set to predict each model's accuracy and error rate.

4.10 Summary

This chapter concluded the methodology of this PhD thesis experimental study. Ethical approval has been explained deeply. Raw data has been collected, prepared, and pre-processed by data cleaning, detecting outliers, filling missing values, missing values mechanism, data integration, normalisation, and feature selection methodology. Finally, this chapter discussed the experimental setup of machine learning approaches using single and combined classifiers.

Chapter 5 will provide more details on the classifiers and their results than evaluating them according to the outputs.

Chapter 5 Results and Discussion

5.1 Introduction

This chapter discusses the simulation results and analysis of the student level classification. There are two main sections presented in this chapter. First, two single classifiers were used to evaluate the proposed models in more depth by utilising the standard performance measurement metrics discussed in Chapter 4: Sensitivity, Specificity, Precision, J1, F1 score, and Confusion Matrix, Accuracy, AUC, and ROC. Machine learning classifiers provide significant properties, such as non-linear mapping, universal approximation, and parallel processing [138]. Second, single classifiers are combined with different features to produce a more effective model and provide better results. These models are demonstrated as a crucial procedure for many applications, including the education field.

5.2 Single Machine Learning Classifiers Results for Classification

This section demonstrates the classification outcome for student levelling datasets from different schools. The classification outcome has been achieved using the feature selection method on 24 features from the student dataset collected from two schools using UK and US systems. After implementing the feature selection method, the 15 features have been chosen as having the most decisive influence on the class label.

A dataset is applied to some selected models. The dataset is split into training (70%), validation (10%), and testing sets (20%). The training set is a part of the dataset that the model uses to learn from to operate correlational tasks (weights and biases for NN algorithm). The validation set is applied during the training process (a small portion of the training set) provides an unbiased evaluation of the training set by tuning the model's hyperparameter. The testing set has a different purpose; it assesses the performance of classifiers with unknown class labels. Having multiple datasets allows benchmarking the performance evaluation metrics from all those datasets [139].

5.2.1 Random Forest Classifier (RFC)

The performance evaluation techniques were achieved using the collected student levelling dataset of 1550 samples. The imperial study in RFC was performed using random forest models. The classification performance was evaluated using the evaluation metrics discussed in previous chapters. During simulation, both training and testing datasets were selected randomly whilst repeating in every test run. The RFC models were applied to the 16 features (including the class label). The results gained from this experiment produced reasonable values, as shown in Table 5.1. Training and testing stages were applied to the dataset and measured by the performance measurements as shown in the below tables. The proposed method's activity has also been evaluated in visual performance evaluation with ROC and AUC charts, as shown in Figure 5.1 and Figure 5.2.

The RFC model was built and trained using multiple trees prepared using the student levelling dataset; 50, 100, 200, 400, 500 and 1000 trees. RFC100 had generated the best Accuracy result during the training process with an average of 0.69 for the three classes, but the AUC average was 0.68, the third-highest result. Although the AUC was higher using the RFC50 method (0.72), it had a lower 0.66 Accuracy than the RFC100. The Sensitivity of RFC200 performed better than the other methods with 0.661. As shown in Figure 5.1, the RFC50 demonstrates the best results of the ROC curve among all other models.

Table 5.1 RFC average performance of the 3 classes (Training sets)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
RFC50/1	0.632	**0.674**	**0.466**	**0.473**	**0.306**	0.66	**0.72**
RFC100/2	0.592	0.673	0.434	0.460	0.265	**0.69**	0.68
RFC200/3	**0.661**	0.617	0.416	0.443	0.278	0.63	0.69
RFC400/4	0.524	0.657	0.405	0.416	0.181	0.65	0.61
RFC500/5	0.579	0.585	0.371	0.404	0.163	0.61	0.61
RFC1000/5	0.465	0.553	0.344	0.350	0.019	0.541	0.514

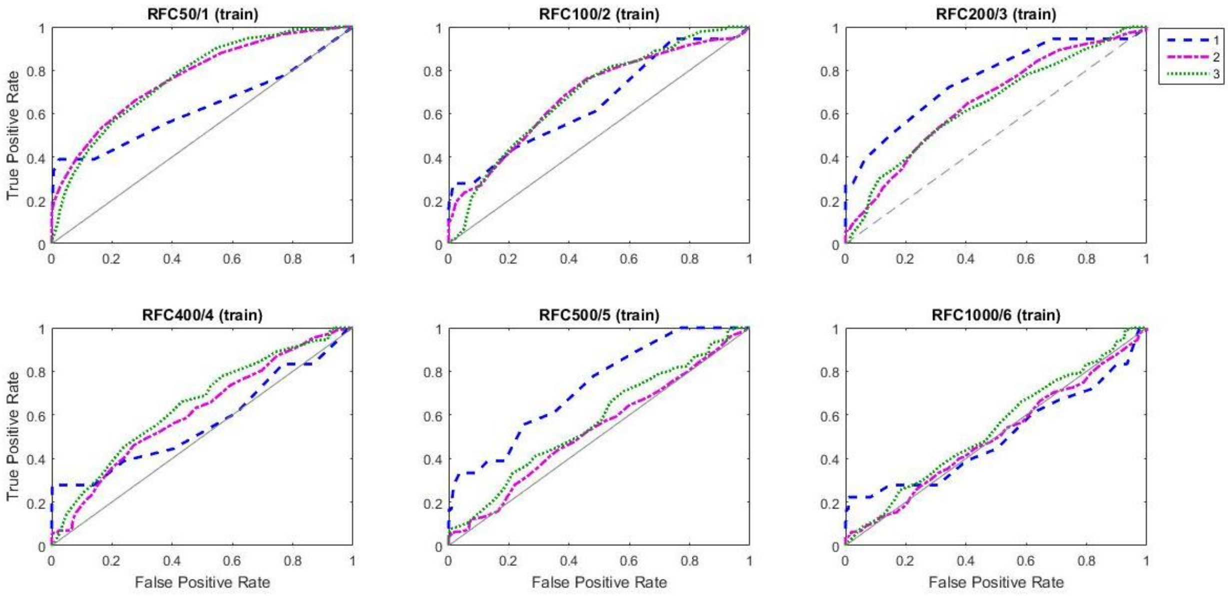


Figure 5.1 ROC Curve Training results for the RFC

Random forest classification combines the basics of decision trees with additional flexible parameters, which increases the model's Accuracy [140]. The bootstrapped method allows multiple times a selection of critical samples. Once the bootstrapped datasets are made, a random subset of variables is used to develop the random forest [141]. Fifteen input features and the class label from the students' dataset are considered for every step. Viewing the subset of variables for each step, a new bootstrapped dataset is developed alongside several RF trees, and this process was completed and repeated several times. Once the updated students' dataset is ingested into all the trees in RF, a calculation is done to discover which model received the highest number of votes.

In the training stage of the model, RFC50 had the highest results among all the RFC models, for classes 2 and 3, as shown in Figure 5.2. However, the average of classes produced a specificity of 0.674, Precision 0.466, F1 score of 0.473, j score 0.306, and AUC 0.72. The AUC for RFC100 and RFC200 are 0.68 and 0.69, respectively, while RFC1000 has the lowest AUC performance.

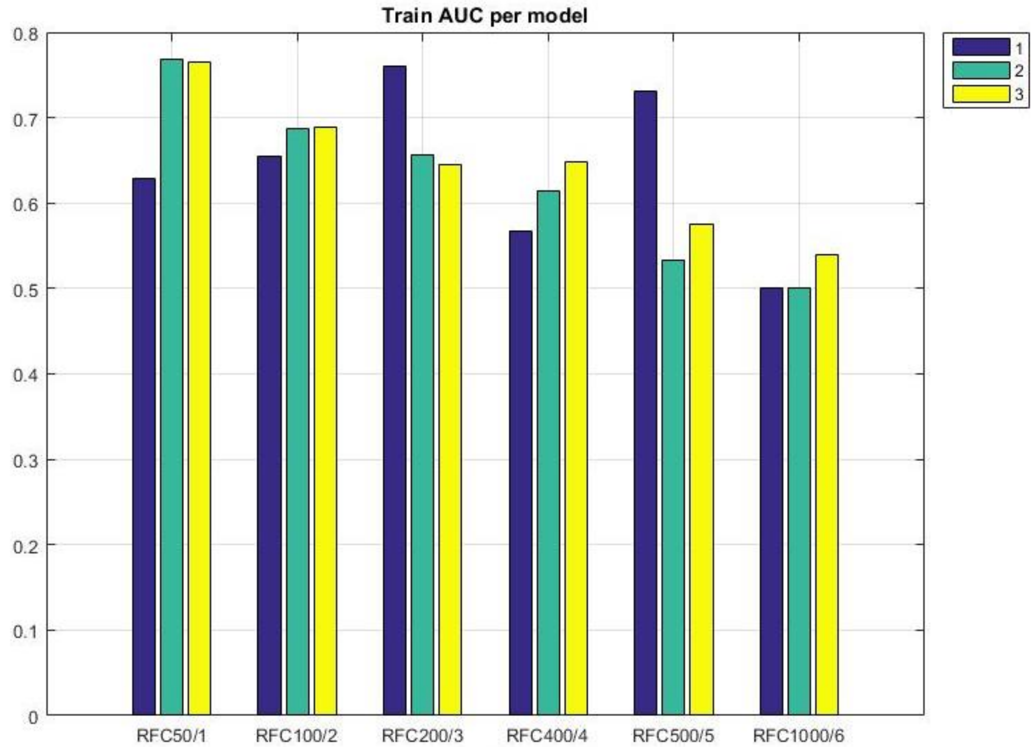


Figure 5.2 AUC Training for RFC

The RFC models' performance during the testing phase is slightly lower than the results during the training stage. As shown in Table 5.2, the RFC50 generated the highest AUC with an average of 0.587, whereas RFC100 generated the highest Accuracy of 0.634. On the other hand, RFC200 generated the lowest AUC compared to all other RFC models. RFC50 yielded the most heightened Sensitivity of 0.709 too, whilst RFC500 scored 0.621 to be the second-highest one in terms of Sensitivity. RFC50 had the second-lowest specificity of 0.483, whilst RFC100 generated the highest specificity of 0.656.

As shown in Figure 5.4, the RFC50 has given the best performance of ROC in testing and training stages as per the AUC readings. Figure 5.1 and Figure 5.3 show the visual representations of the ROC curve for training and testing stages, respectively.

Table 5.2 RFC models for the testing samples

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
RFC50/1	**0.709**	0.483	**0.417**	**0.439**	0.192	0.505	**0.587**
RFC100/2	0.493	**0.656**	0.388	0.393	0.149	**0.634**	0.554
RFC200/3	0.422	0.608	0.393	0.379	**0.196**	0.573	0.463
RFC400/4	0.508	0.55	0.368	0.381	0.03	0.556	0.527
RFC500/5	0.621	0.446	0.35	0.395	-0.52	0.515	0.531
RFC1000/5	0.616	0.332	0.325	0.347	-0.052	0.361	0.389

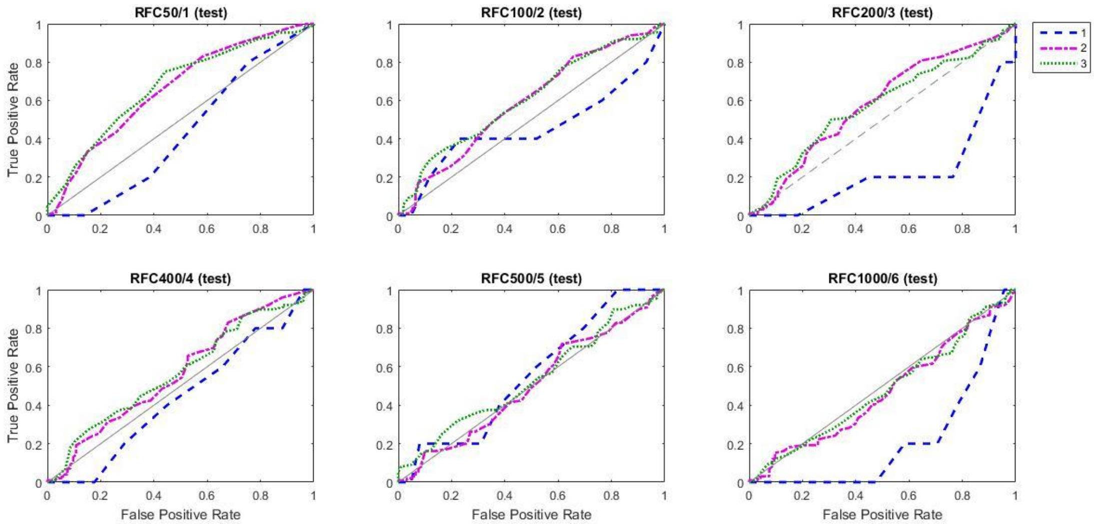


Figure 5.3 ROC curve Testing for RFC Models

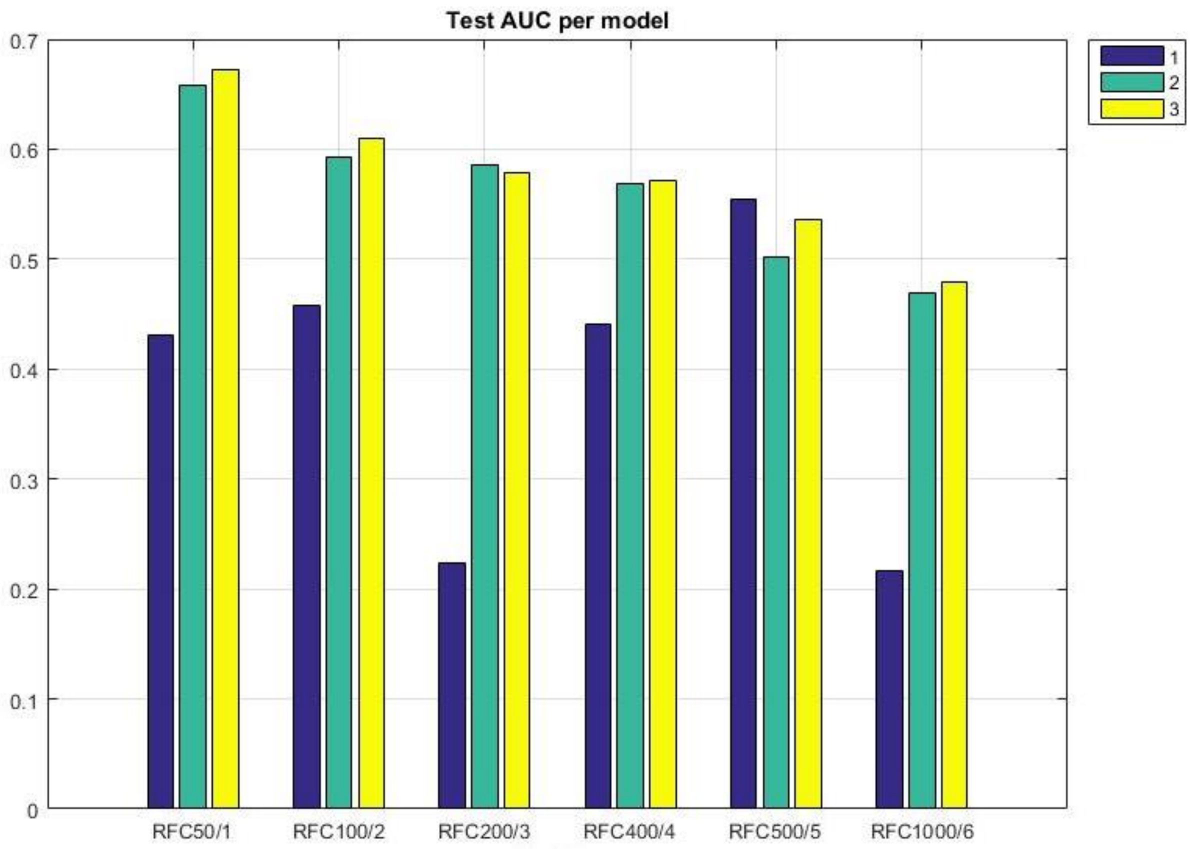


Figure 5.4 AUC (Testing) for RFC

As implied from Figure 5.4, results of class 1 for all the RFC models at the testing stage are fluctuated highly between 2.2 and 5.5 using the AUC performance measurement, but the best result were on the 500 trees model at 5.5. The Second and Third classes had shown close results to each other amongst all these models and the highest results were found at the 50 trees model. The three classes average best results of 0.7 were found at the 50 trees model.

5.2.2 Artificial Neural Network

The ANN can perform multiple classifications, clustering, and dimensionality reduction [62]. The purpose of implementing NN is to measure the network with generalisation capability using the weigh unit throughout the construction of the NN network and compare the model's performance with other classifiers.

The evaluation of the model is achieved using the holdout method implemented on 1550 samples. The model received three sets of datasets: 70% for training, 20% for testing, and 10% for validation, as shown in Table 5.3. This research utilised two methods of NN models: Levenberg Neural Network Classifier (LEVNN) and Backpropagation Network Classifier

(BPXNC). Although LEVNN obtained a higher accuracy average than BPXNC, they both generated the same AUC value of 0.962, as shown in Table 5.4. Overall, all the parameters have shown that LEVNN caused better results than BPXNC.

Table 5.3 Dataset samples

Dataset	Number of Samples
Training	1085
Testing	310
Validation	155

Table 5.4 ANN average performance of the three classes (Training sets)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
LEVNN	**0.952**	**0.929**	**0.697**	**0.73**	**0.881**	**0.936**	**0.962**
BPXNC	0.935	0.92	0.692	0.718	0.855	0.925	**0.962**

Figure 5.5 ROC curve Training sets for LEVNN and BPXNC models show the ROC curve visualisation at the training stage of LEVNN and BPXNC, which demonstrates that these models performed well and equal at this stage; the area under the curve in both models are the same size. Figure 5.6 shows the AUC of the average of the three classes for LEVNN and BPXNC.

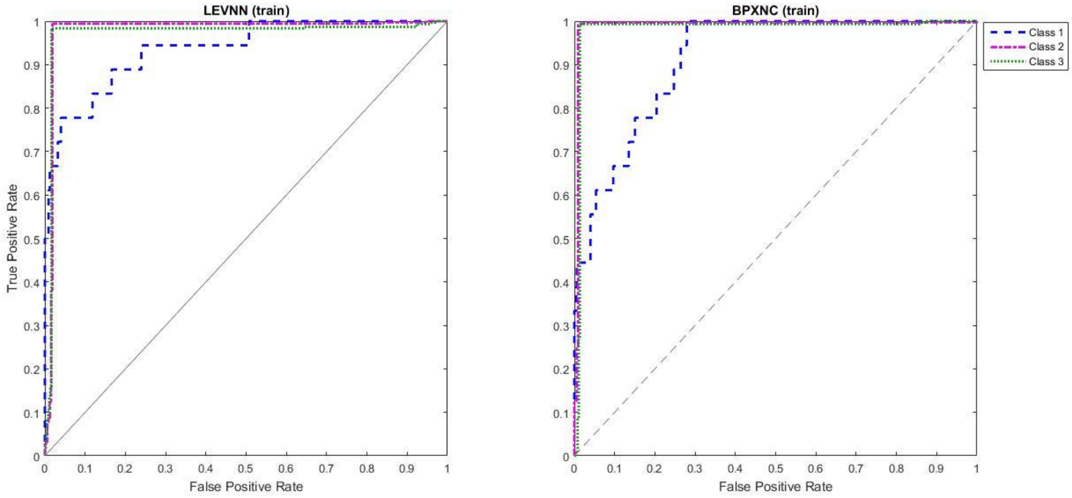


Figure 5.5 ROC curve Training sets for LEVNN and BPXNC models

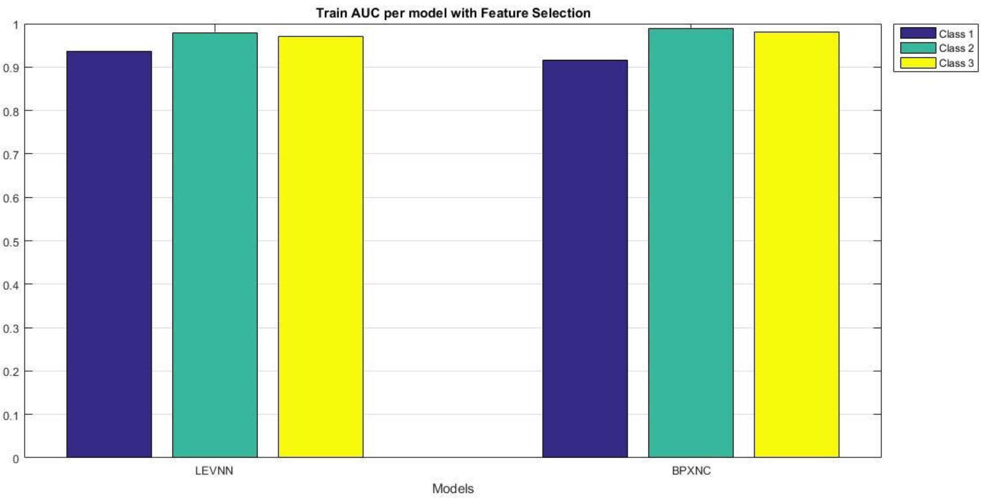


Figure 5.6 AUC Training sets for LEVNN and BPXNC

Despite receiving high accuracy results during training for both LEVNN and BPXNC, the testing results performance for both NN models was slightly lower than the training process. For LEVNN, the hidden layers were amended using backpropagation links, whereas BPXNC were altered using feedback from the output layer. Table 5.5 shows the classification performance evaluation for LEVNN and BPXNC during testing. Although BPXNC demonstrates a slightly better performance of Precision (0.571) and AUC (0.767) than the LEVNN model, generally, LEVNN has produced higher scores for most of the parameters with 0.801, 0.815, 0.602, 0.616, and 0.816 for the Sensitivity, Specificity, F1, J, and Accuracy, respectively. Figure 5.7 shows the ROC curve illustration of LEVNN and BPXNC, implying that these models perform well but slightly better performance of the BPXNC on 0.767 over the LEVNN result of 0.732 AUC, the area under the curve is slightly bigger in BPXNC. Figure 5.8 shows the AUC of an average of the three classes for LEVNN and BPXNC with the distinction of the second model.

Table 5.5 ANN performance (Testing sets)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
LEVNN	**0.801**	**0.815**	0.568	**0.602**	**0.616**	**0.816**	0.732
BPXNC	0.792	0.789	**0.571**	0.577	0.581	0.774	**0.767**

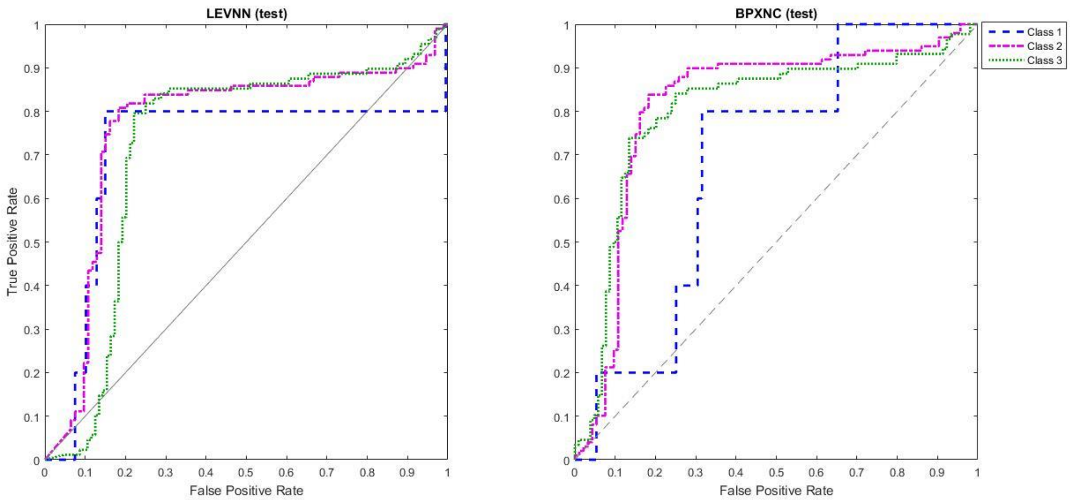


Figure 5.7 ROC curve for ANN (Testing sets)

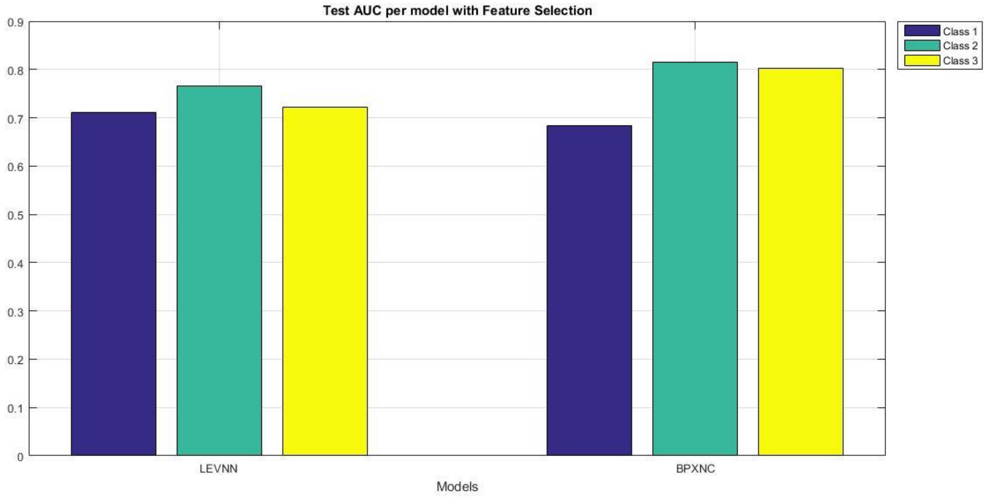


Figure 5.8 AUC plot for ANN (Testing sets)

5.3 Combined Classifier

A combined classifier is a method that combines two or more classifiers to improve the performance and accuracy of the models [132]. This model of machine learning algorithms is found to generate good results when combining specific classifiers. Implementing a combined classification in this study allowed the models to learn from the non-linear components and yield good results. The combined classifiers use a pattern recognition system combined with a bootstrap aggregating approach to enhance the selected models [142].

The proposed study implemented several combined classifiers, including NN Com, LEVNN Com, and NN and RF Com1 and Com2. The classification performance evaluation method is based on 16 elements and three classes of student levelling datasets. Table 5.6 shows the outcome for combined classifiers in training. The prediction of the student levelling dataset during the training phase is achieved by taking the majority vote of RFC for several cycles. Alternatively, NN generated the predictions after several processes by taking a weighted vote.

Table 5.6 Combined classifiers of the three classes' performance average (Training sets)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
NN Com	0.997	0.995	1	0.997	0.992	1	1
LEVNN Com	0.997	0.995	1	0.997	0.992	1	1
NN and RF Com 1	0.999	0.999	0.999	0.999	0.998	0.999	1
NN and RF Com 2	0.999	0.999	0.999	0.999	0.998	0.999	1

The overall AUC for all the combined classifiers were equal to one during the training process in this study. Although the Precision for NN Com and LEVNN Com was one compared to 0.999 for NN and RF Com1 and Com2, the overall performance for NN combined with RF was still performing better (the average of all the parameters for NN and RF Com1 and Com2 was 0.999, but for NN Com and LEVNN Com was 0.997). A good combination between NN and RF is highlighted, indicating that vital information is available in the selected datasets. Figure 5.9 shows the ROC curve illustration of the combined classifiers used the research, implying that all these models performed very well at the training stage. Figure 5.10 shows the AUC of the three classes the combined classifiers at the training stage, which indicates perfect results of 1 for all of them at this stage.

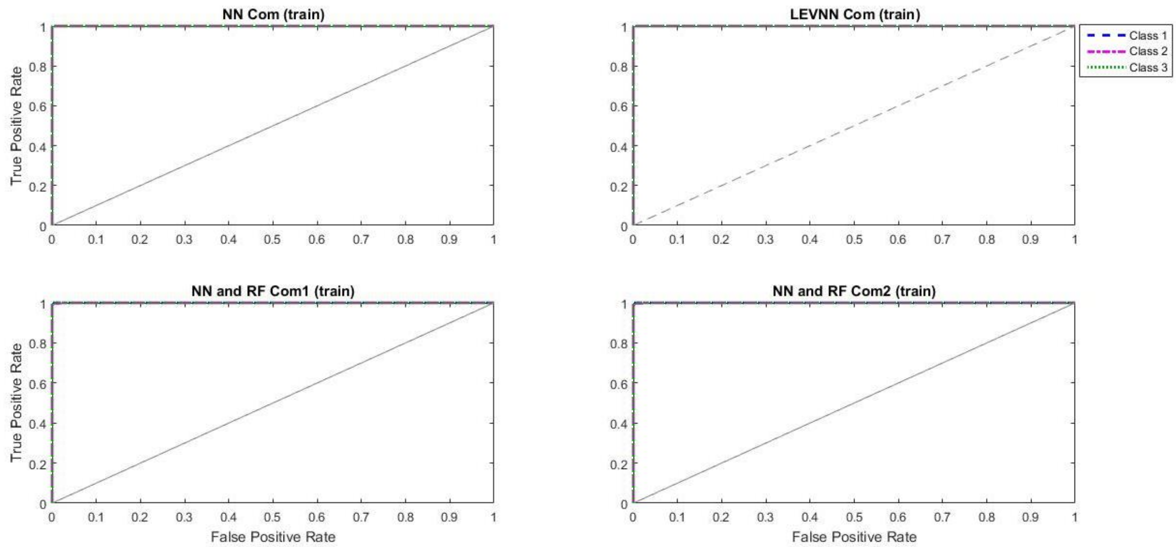


Figure 5.9 ROC curve for the combined classifiers (Training sets)

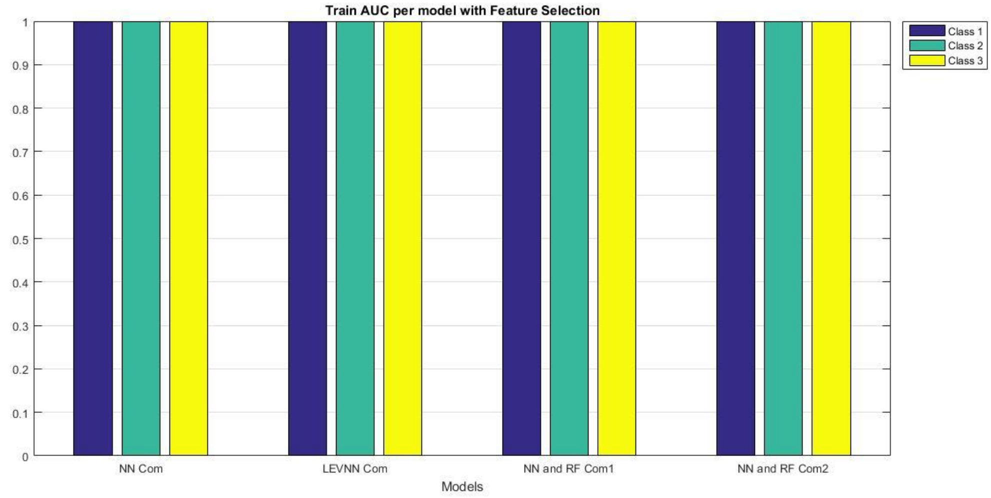


Figure 5.10 AUC plot for the combined classifiers (Training sets)

The algorithm's performance was way lower at the testing stage than the training of combined classifiers; NN and RF Com1 obtained a Sensitivity of 0.755, Specificity 0.798, Precision 0.565, F1 0.585, J Score 0.552, Accuracy 0.799, and AUC 0.824, which is the best results compared to all the combined classifiers including the NN and RF Com2 that show the best sensitivity reading with slight differences, as shown in Table 5.7 Combined classifiers performance (Testing sets).

Figure 5.12 shows the ROC curve illustration of the combined classifiers used the research, implying that all these models performed very well which means a perfect results for the testing stage. Figure 5.11 shows the AUC of the three classes results of the combined classifiers in the testing stage, showing the best result for NN and RF Com 1 at this stage with the biggest size of the area under the curve for all the classes.

Table 5.7 Combined classifiers performance (Testing sets)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
NN Com	0.769	0.76	0.538	0.549	0.529	0.748	0.712
LEVNN Com	0.773	0.656	0.528	0.533	0.429	0.649	0.663
NN and RF Com 1	0.755	**0.798**	**0.565**	**0.585**	**0.552**	**0.799**	**0.824**
NN and RF Com 2	**0.776**	0.722	0.542	0.547	0.498	0.712	0.787

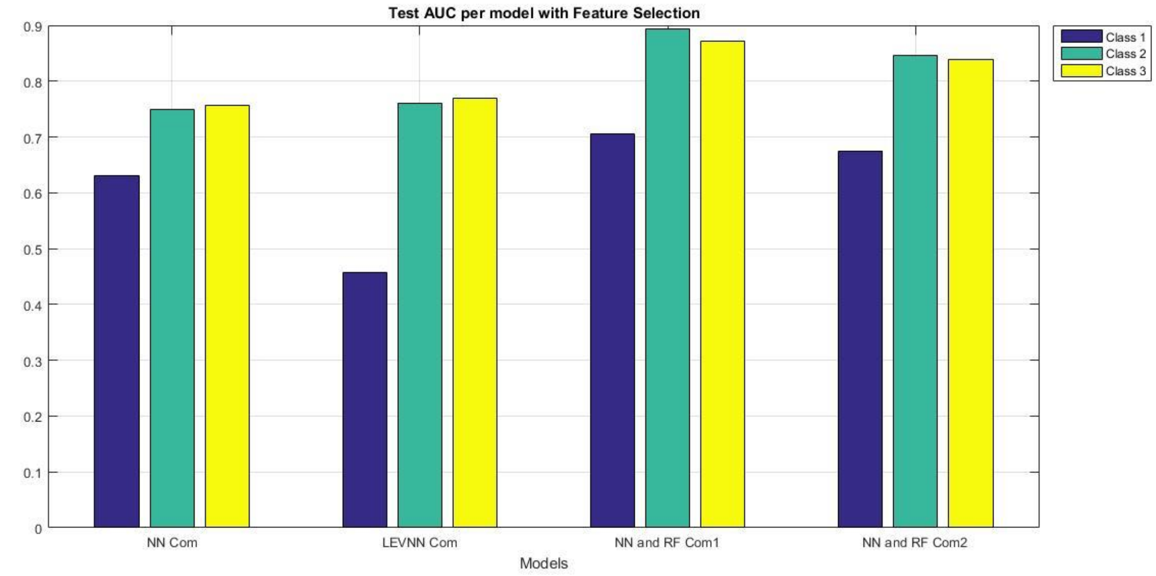


Figure 5.11 AUC for the combined classifiers (Testing sets)

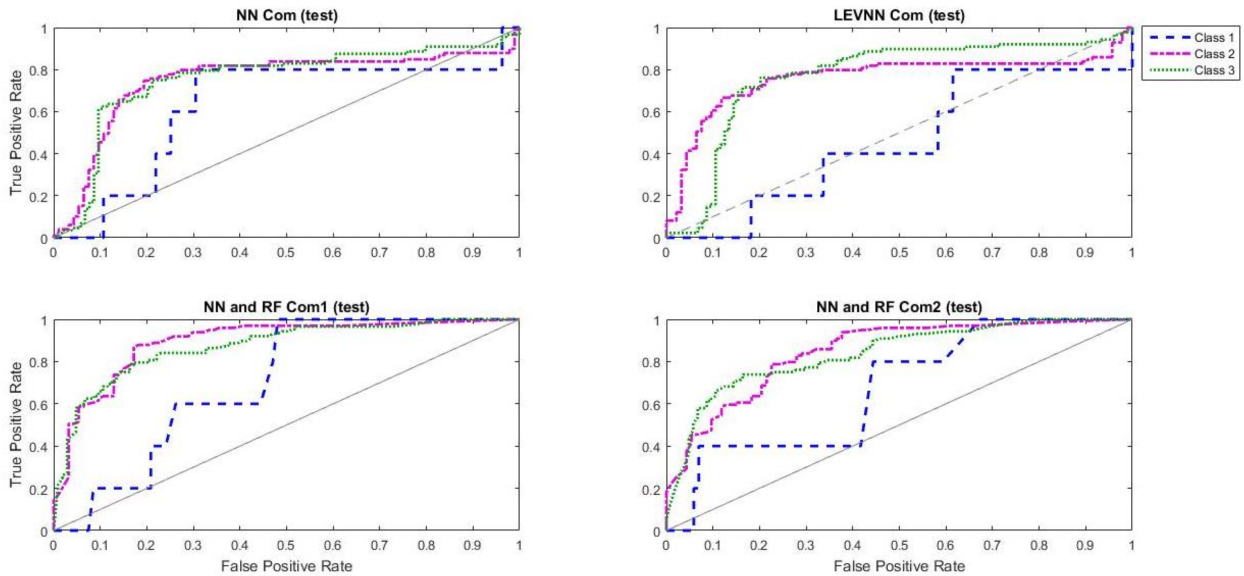


Figure 5.12 ROC curve for the combined classifier (Testing sets)

5.4 Discussion

In this research, a data science methodology combines 16 features extracted from 1550 records to predict the student levelling. The chosen student dataset demonstrates non-leaner relationships, which creates a real challenge to the classifiers. The RFC classifiers performance is poor compared to the other classifiers, showing that RFC does not have the capabilities to handle the training data and unseen examples. However, the NN and RF Com1 performed well during the training stage, generating the highest performance. Yet, it did not perform as well during the testing process as the results were way lower than the training ones, but it generated the highest AUC of 0.824 over the rest of the classifiers.

LEVNN model produced the best performance results amongst all the models applied in this research during the testing process except the Precision and AUC. The highest figures of Sensitivity, Specificity, F1, J, and Accuracy are 0.801, 0.815, 0.602, 0.616, and 0.816, respectively. In contrast, the best results of AUC were achieved by NN and RF Com1 of 0.824. The AUC for the combined classifier (NN Com, LEVNN Com, NN and RF Com) generated 1 for the training sets, while testing generated AUC of 0.824 for the NN and RF COM1, which is the best result of all the algorithms used in this study. Compared with the other tested machine learning models, RFC did not show high Accuracy results.

Because the LEVNN method achieved the best performance outcomes during the testing phases, it has been considered the best classifier in this research, which is more important than training.

Overall, the type of results gained highlights the potential of Multi-National Schools' data for classifying the student levelling. The choice of model is essential when needing accepted results, as shown in the classification performance tables for training and testing in this research. The LEVNN classifier responded well to the students' data and has potential use in education.

The LEVNN is a powerful model for analysing the students' datasets, and it has proven in this domain that it presented substantial prediction accuracy and performance compared to other classifiers. A good relationship between input features and target values is discovered during the development process. The datasets are moderate in size, with 20% of the input features randomly selected for testing and the remaining percentages of 70% and 10% used for training and validation, respectively. In this context, the test set errors are averaged, and the procedure was repeated several times.

Sikder et al. [143] also used LEVNN in their study for predicting 120 students' performance with a training dataset of 70% from the total dataset, a testing dataset of 15%, and a validation dataset of 15%. They used the following features: Class Test Mark, Family Education, Class Performance, Living Area, Home, Class Attendance, Social Media Interaction, Assignment, Extra-Curricular Activity, Lab Performance, Drug Addiction, Study Time, Affair, Previous

Result, and Year Final Result, living place and social interactions. In case of the success of their study, they will prove that students' performance can be affected by their lifestyle. Their study has shown good accuracy results for both outstanding and poor GPA performance of students with better accuracy for the poor performance giving an average of the accuracy of 97.2%. They claimed that a previous study used the ANN algorithm to predict students' performance using multilayer perception trained by static backpropagation. It produced less precision than the first study, with an average of 74%.

Although Table 5.8 shows that the ensembles classifiers have the best results among all other algorithms, LEVNN is provided the best results among the other models used in the research at the testing stage for predicting students' performance for levelling usage, given a dataset of previously mentioned attributes, as shown in Table 5.9. It produced good prediction accuracy and performance compared to the other classifiers.

Table 5.8 Classification performance (Training stage)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
RFC50/1	0.632	0.674	0.466	0.473	0.306	0.66	0.72
RFC100/2	0.592	0.673	0.434	0.460	0.265	0.69	0.68
RFC200/3	0.661	0.617	0.416	0.443	0.278	0.63	0.69
RFC400/4	0.524	0.657	0.405	0.416	0.181	0.65	0.61
RFC500/5	0.579	0.585	0.371	0.404	0.163	0.61	0.61
LEVNN	0.952	0.929	0.697	0.73	0.881	0.936	0.962
BPXNC	0.935	0.92	0.692	0.718	0.855	0.925	0.962
NN Com	0.997	0.995	**1**	0.997	0.992	**1**	**1**
LEVNN Com	0.997	0.995	**1**	0.997	0.992	**1**	**1**
NN and RF Com 1	**0.999**	**0.999**	0.999	**0.999**	**0.998**	0.999	**1**
NN and RF Com 2	**0.999**	**0.999**	0.999	**0.999**	**0.998**	0.999	**1**

Table 5.9 Classification performance (Testing stage)

Model	Sensitivity	Specificity	Precision	F1	J	Accuracy	AUC
RFC50/1	0.709	0.483	0.417	0.439	0.192	0.505	0.587
RFC100/2	0.493	0.656	0.388	0.393	0.149	0.634	0.554
RFC200/3	0.422	0.608	0.393	0.379	0.196	0.573	0.463
RFC400/4	0.508	0.55	0.368	0.381	0.03	0.556	0.527
RFC500/5	0.621	0.446	0.35	0.395	-0.52	0.515	0.531
LEVNN	**0.801**	**0.815**	0.568	**0.602**	**0.616**	**0.816**	0.732
BPXNC	0.792	0.789	**0.571**	0.577	0.581	0.774	0.767
NN Com	0.769	0.76	0.538	0.549	0.529	0.748	0.712
LEVNN Com	0.773	0.656	0.528	0.533	0.429	0.649	0.663
NN and RF Com 1	0.755	0.798	0.565	0.585	0.552	0.799	**0.824**
NN and RF Com 2	0.776	0.722	0.542	0.547	0.498	0.712	0.787

5.5 Chapter Summary

This study performed an empirical investigation into implementing several machine learning models to classify student levelling. This study aimed to investigate the effectiveness of machine learning models when handling student data. It was found that through the experiments conducted, using student data in combined classifiers, LEVNN generated results indicating the goal of predicting the level of the students is viable and yields good and accepted results. The results show that this classifier was the most effective when dealing with students' levelling data.

Chapter 6 Web-Based Application System

6.1 Introduction

Machine learning and rule-based systems are used to make implications from a range of stored data. Both approaches have their strengths and weaknesses; therefore, choosing the appropriate system is crucial. The system that will be developed is a rule-based online interactive system built for schools to manage their students in three different areas, admission, levelling, and differentiation. The detailed information about the web-based student levelling system will be discussed in the following section, which can be accessed from anywhere globally and through any browser. As the system is planned to be built using the bootstrapping framework, the web solution will automatically adjust to the screen sizes of the devices accessed from [144]. This Solution helps save the complete information of students whilst analysing the student background of education, thereby proposing their eligibility, predicting the students' performance and being a robust platform allowing further development by adding new modules to the system. The Cloud web application is developed module-based using HTML, CSS, JavaScript, and PHP on a virtual machine based on CentOS Linux Machine.

6.2 System Requirement Specifications

The design techniques needed to develop the system is a data-based construction that follows a machine learning approach. The data need to reflect important patterns by using the modelling techniques. The model can be revised automatically as the database in the cloud is updated with the received data from schools. Their decisions are made regarding the outcome of the student in the admission stage, levelling, and differentiation, and after this sequence of data collection is made enough, the system will be able to make predictions based on the data collected and what it has learnt over time. Finding the first triggering rule by searching in the ruleset (IF-Then rules) is the main aim in the prediction stage. A suitable structure will be needed to represent a rule set adequately; the system will include decision trees and linear lists. A decision tree will consist of a root, internal nodes representing attributes, leaf nodes representing classification, and branches representing attribute values. The linear list describes the rules of ‘IF-THEN’ [145]. The system consists of four main components: knowledge base, inference engine, temporary working memory, and user interface.

The Student Management Systems need to utilise IoT devices, fog computing and cloud computing and establish a seamless data exchange between schools’ admission and their teachers. The system Provides protocols to support the communication and broadcast of raw student data from affiliated schools and smart devices to a network of fog nodes. The corresponding school data will need to be stored at two different storage layers: the fog and cloud. The user will have the ability to access stored data in the cloud for the respective school; nevertheless, once a student has transferred from school to school, that data will be shared mutually until the student is registered into the new school. Analysis of collected student data in terms of levelling should have accessibility within the organisation and from any location and have the privileges to read and write the data. With the use of stored data in the cloud, teachers will monitor their students remotely and provide timely feedback plans to be held in the fog to work on students’ strengths and weaknesses. The use of cloud computing and fog computing will enable schools to share data in real-time processing, improve data privacy, keep track of student admissions and levels, and minimise gaps in education.

The approach combines the best strengths and synergies of cloud computing and machine learning technologies to effectively analyse the data and develop predictive analysis capabilities, actionable information, better student levelling information, and decision making. Technically, by combining and leveraging cloud computing and ML technologies, our primary goals of the suggested framework are included (not limited to):

1. Sustaining better and effective admissions policies.
2. Deciding the year/grade group of students.
3. Delivering a smooth transition for students when transferring from school to school with the same or different curricula.
4. Continuously generate the student’s level and provide essential levelling criteria for teachers to follow.
5. Differentiate students by providing information on areas that show signs of weakness and methods to provide individual student learning.

The main goal of the suggested framework is to realise the requirements by analysing the data and transforming information into knowledge. Requirements for the system is to find valuable insights, patterns and trends in data that can lead to actionable information, decision making, prediction, situation awareness and understanding. To complete those technical tasks required by the system, the developed framework need to leverage machine learning algorithms, knowledge mining, and knowledge-intensive problem-solving.

This research's general scenario features a constant interplay between the cloud layer, fog layer, and physical sources (schools). The cloud layer will need to oversee storing all the prediction data, students' historical backup data, and heavy storage operation in general. The cloud will be the leading player in the workings of the framework due to its ability to concentrate powerful data centres [146]. The cloud will be the significant data stream to the fog layer and intra-layer and inter-layer communication. Both the cloud and the fog will be part of the communication with the final user.

6.2.1 System Architecture

In the following section, the general framework architecture will be described, then the cloud and fog layer integration as signified in Figure 6.1. The framework consists of three layers; intelligent decisions (Stakeholders), fog layer and cloud layer. Schools make smart decisions using different devices (PC, Laptop, Smartphone, and Tablet) to easily send various student data and requests through cloud computing to obtain other choices and levelling reports. Each network has several application hosts = (H_1, H_2, H_n) providing the Software as a service (SaaS) and can be allocated to execute the cloud stakeholders that make the intelligent decisions [147]. Each application host has a set of resources = (R_1, R_2 and R_n) allocated for school requests. Each network has a network administrator responsible for coordinating the communication between the hosts inside the networks and other networks in the cloud. A network administrator is responsible for running the algorithm.

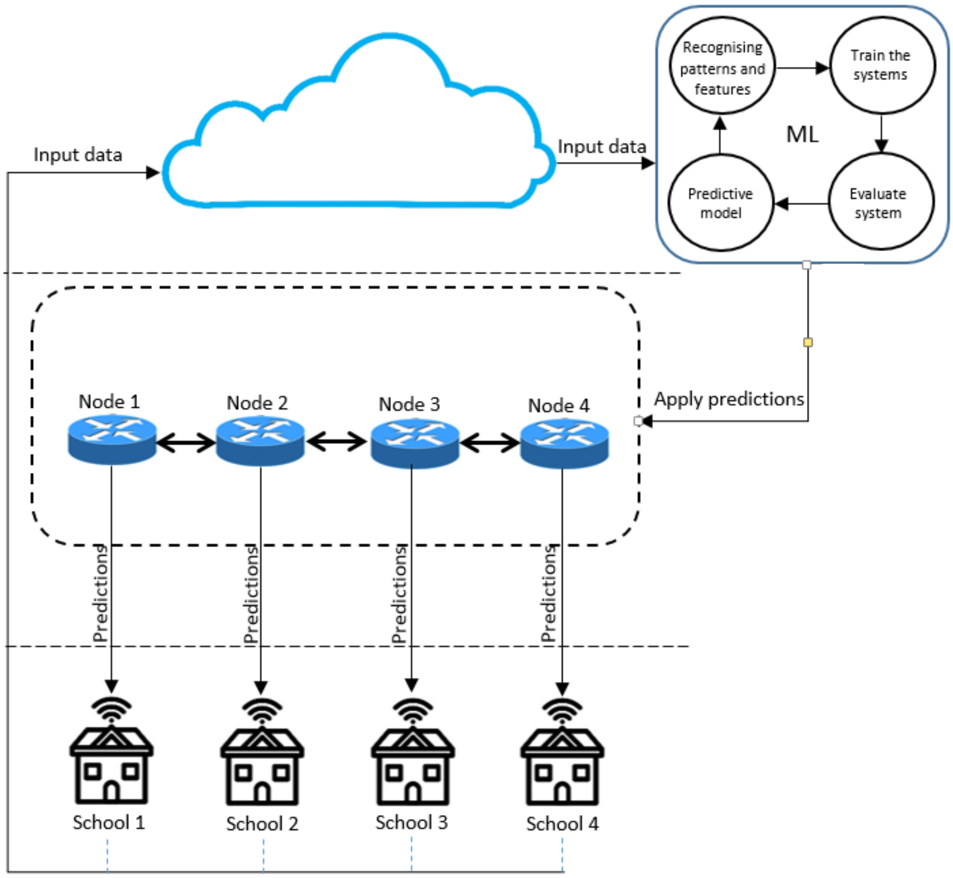


Figure 6.1 Proposed system framework overview

As shown in Figure 6.1, each school will have access to the system, and data moved to and from the cloud will be stored there. The cloud is a virtual machine and then further split into small virtual machines separated by each port [146]. Amazon server being the first server to interact with the user, it will act as the fog layer, and then it is connected to the database. Finally, the collected dataset will be passed onto the ML tools. The data flow among servers will be controlled via web application of the Amazon server and thereby to the database.

6.2.2 Front-End and Back-End System

X data is inputted to the browsers to visualise a real-time scenario. The routing and controllers will pass the student's information to the modals and store it in the database. Once the student data is subjected to levelling, the request is sent to the server again. The data gets downloaded back to the school nodes and the brain.js, which is now subjected to the pre-defined algorithm of decision making, which will push out the students' results and await the user’s confirmation for levelling. Once the proof is received, depending on the level of access to the school, the data gets sent back to the routes, controllers, modals and finally stored back in the database as different datasets. The dataset stored in the database is connected to the machine learning algorithm in the TIBCO tools (subjected to program the ML). Once the ML is programmed with the preferred algorithms, we will process the datasets with different parameters and train the algorithm with the training dataset.

6.2.2.1 Accessing the system

The levelling system will need to be accessed using any web browser with the shared link. Once the link is accessed online, the user will be directed to a web-based dashboard interface (UI). The involved schools have the login credentials to access the system via the login screen. The system will need to work on two types of logins, 'School System Admin' and 'User'. The School System Admin will manage the users in the school and control what kind of access the users will have. The login credentials for the system will need to be in the form of a "Username" and "Password".

Once the users have signed in, they will need to be directed to a home screen where the options available will be "My Profile", where the user can personalise the setting as per the requirement/work profile. In addition to "My Profile", there will be an indicator that reflects two colours, 'Red' or 'Green', a red indicates that the user has not wholly logged in and the system is not live. However, green indicates that the user has successfully logged in.

'My profile' will show the personalisation options that the user can add to the home screen, which will include:

- Student Management
- Student Levelling
- Notification Centre
- Report Manager

As there will be several users capable of accessing the system based on the user work profile, the options must have the ability to be configured for generating the home screen as per the requirement of the user. Set up can be saved by selecting "Save Settings" for all future usage, so the users do not have to configure it every time they log in. An icon will need to be provided for the user to view and access all the management options, as shown in Figure 6.2 Widgets

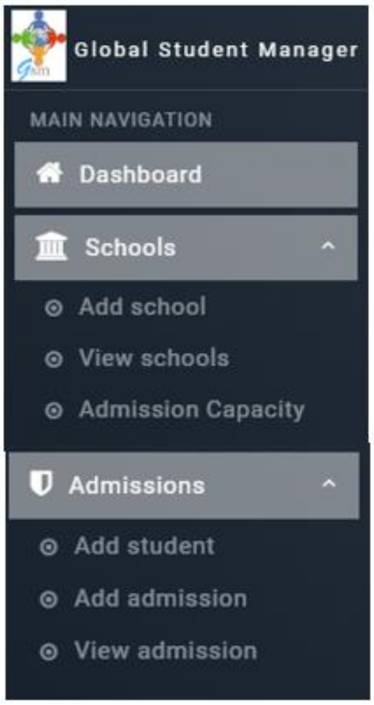


Figure 6.2 Widgets

The "Student Management" action will generate three sub-options to select, which are "Admissions", "Student Holding Pool", and "Student Details".

Admissions – Loads up input types, "Pending Admissions" and "New Admissions". Pending entries can be loaded by searching for the student ID in numeric form or by searching names in text form. As a result, the progress of the student's admission process and the pending actions will be generated. In contrast, the New Admissions will be in the form of an action button that will load the "Admission Form", where all the student's information will be inserted.

As shown in Figure 6.3, the "Admission Form" will include the below input of data:

- Student ID = Numeric Input
- First Name = Text Input
- Middle Name = Text Input
- Family Name = Text Input
- Gender = Select box input with two options, 'Male' or 'Female.'
- Date of Birth = Date input by year, month, and day
- Previous year/Grade Group = The data will be in the form of select box input ranging from 'FS1-Pre-K, FS2-KG1, Year 1-KG2, Year 2 – Grade 1, Year 3 – Grade 2, Year 4 - Grade 3, Year 5-Grade 4, Year 6 – Grade 5, Year 7 – Grade 6, Year 8 – Grade 7, Year 9 – Grade 8, Year 10 – Grade 9, Year 11 – Grade 10, Year 12 – Grade 11, Year 13 – Grade 12'.
- Previous School = Select box input listing all schools in UAE.
- The previous Curriculum = this will automatically generate output once the 'Previous School' input is selected.

- New School = Select box input listing all schools in UAE.
- New Curriculum = this will automatically generate output once the ‘New School’ input is selected.
- Permanent Home Address = Text Input
- Mobile Number = Numeric Input
- Landline = Numeric input
- Remarks = Text input
- Photo = File input

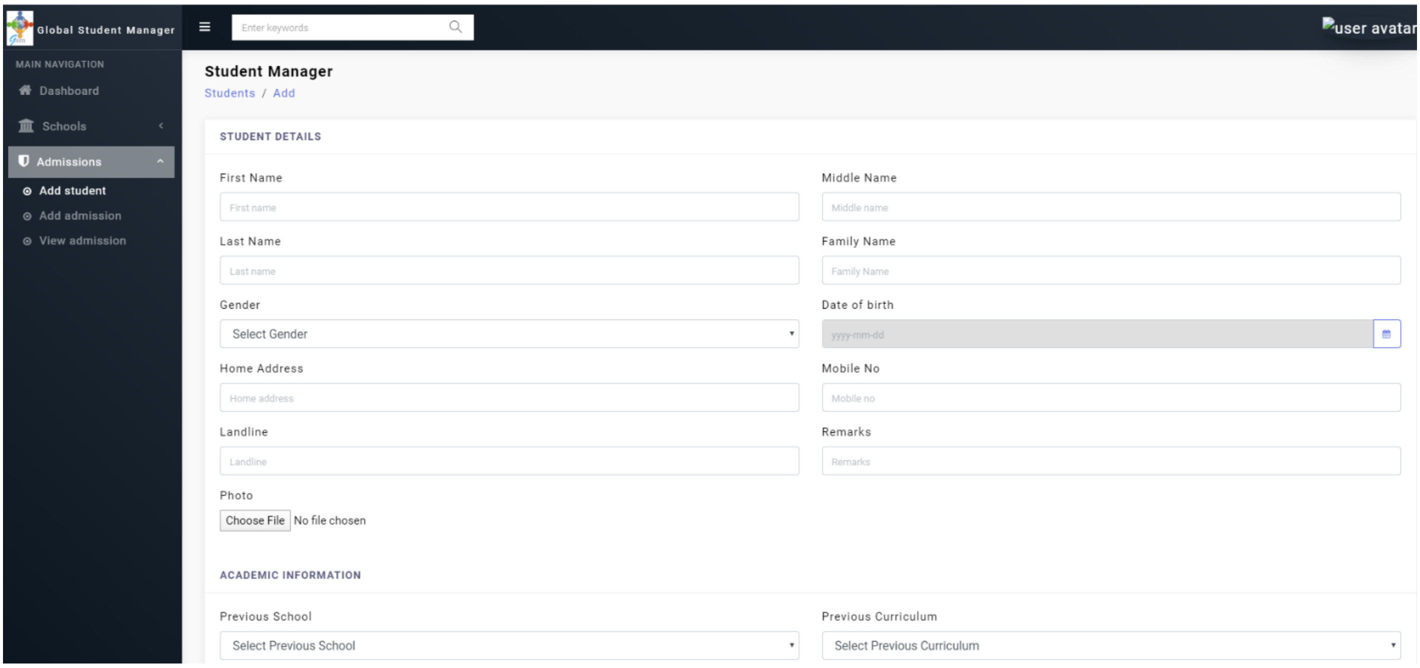


Figure 6.3 Add Students’ window

There are two kinds of assessments that need to be completed during the admission of new students: Interview and ‘Internal Assessment Exam’ as shown in Figure 6.4. The Admission Officer will need to input the interview status, where the input will be in the form of radio buttons inputs ranging from ‘Passed, Partial Pass, or Failed’. The input data for the Internal Assessment Exam will be in the form of radio buttons, which will include ‘Passed’ or ‘Failed’. For every form of data input, there will be an action button “Save” and “Submit”, which will be highlighted once both assessments are complete.

The popup window with “Add Student to Holding Pool” will be triggered depending on the student capacity for that year group. If that particular year group is completed and the student has passed the selection process, the student will be placed in the holding pool. When a student has been given the selection process as per the school requirements and vacancies are available in the year group, the student can be registered directly within that year group. “Assessment Alert” popup will be triggered when the student has failed the Internal Assessment Exam but passed the interview; the student will only be accepted by approving them using the “select box input”. Then, the student will be added to the holding pool. Otherwise, they will automatically be declined. The “Decline Student” popup window will be triggered if the student has failed the interview. All this information will be stored in the cloud for other schools to access. The admission stage will show the year/grade group of the student assigned and the level and differentiation required according to the set logic.

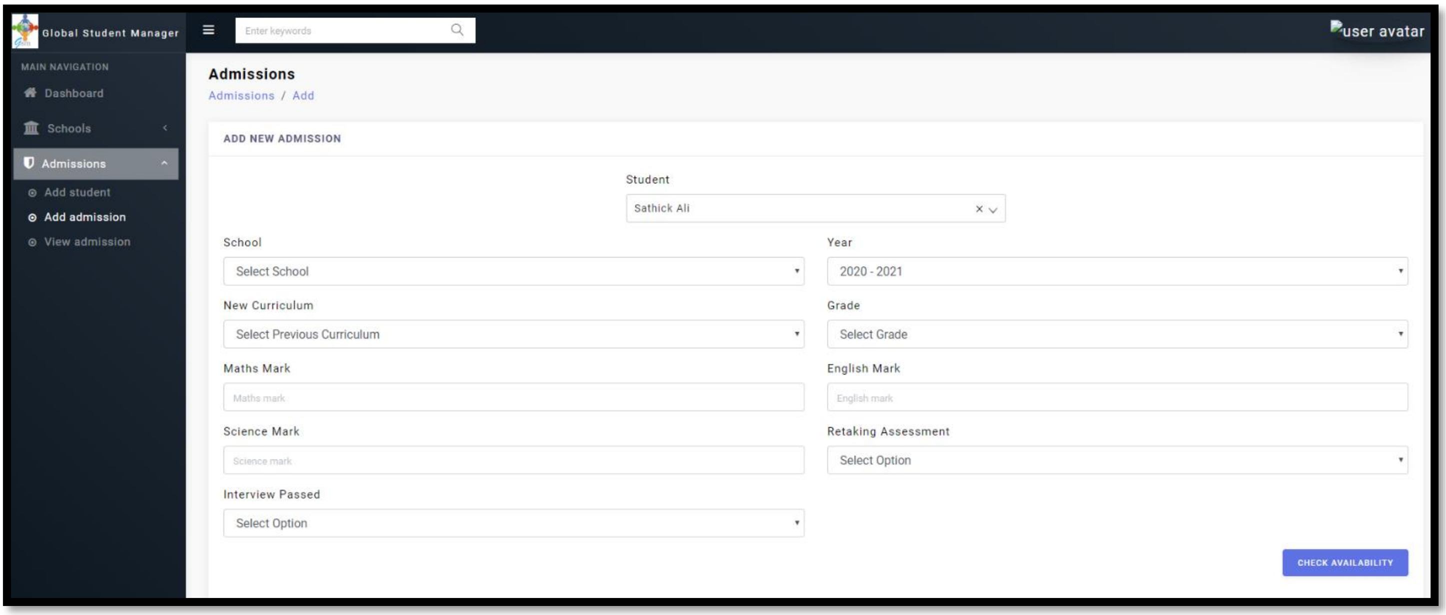
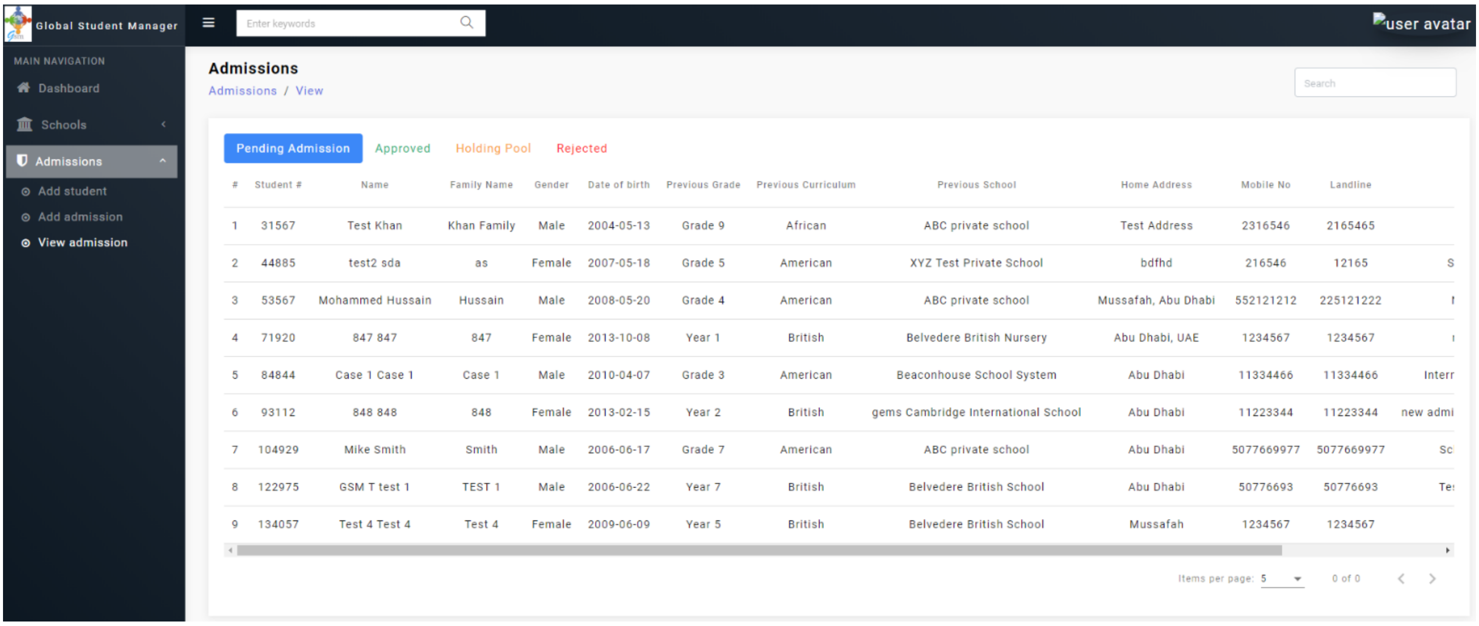


Figure 6.4 Add admissions window

The students within the holding pool can be retrieved by accessing the “Student Holding Pool” page. The user can search for students by ‘Student ID’ in numeric input or ‘Student Name’ in text input as illustrated in Figure 6.4. Once open seats are available within that year group, an alert will be triggered, action is required to be taken. Figure 6.5, Figure 6.6 and Figure 6.7 show the hypothetical statistics of pending admissions, pooling, rejected and approved entries into the school.



#	Student #	Name	Family Name	Gender	Date of birth	Previous Grade	Previous Curriculum	Previous School	Home Address	Mobile No	Landline	
1	31567	Test Khan	Khan Family	Male	2004-05-13	Grade 9	African	ABC private school	Test Address	2316546	2165465	
2	44885	test2 sda	as	Female	2007-05-18	Grade 5	American	XYZ Test Private School	bdfhd	216546	12165	S
3	53567	Mohammed Hussain	Hussain	Male	2008-05-20	Grade 4	American	ABC private school	Mussafah, Abu Dhabi	552121212	225121222	I
4	71920	847 847	847	Female	2013-10-08	Year 1	British	Belvedere British Nursery	Abu Dhabi, UAE	1234567	1234567	ı
5	84844	Case 1 Case 1	Case 1	Male	2010-04-07	Grade 3	American	Beaconhouse School System	Abu Dhabi	11334466	11334466	Interr
6	93112	848 848	848	Female	2013-02-15	Year 2	British	gems Cambridge International School	Abu Dhabi	11223344	11223344	new admi
7	104929	Mike Smith	Smith	Male	2006-06-17	Grade 7	American	ABC private school	Abu Dhabi	5077669977	5077669977	Sc
8	122975	GSM T test 1	TEST 1	Male	2006-06-22	Year 7	British	Belvedere British School	Abu Dhabi	50776693	50776693	Te:
9	134057	Test 4 Test 4	Test 4	Female	2009-06-09	Year 5	British	Belvedere British School	Mussafah	1234567	1234567	

Figure 6.5 View admissions window

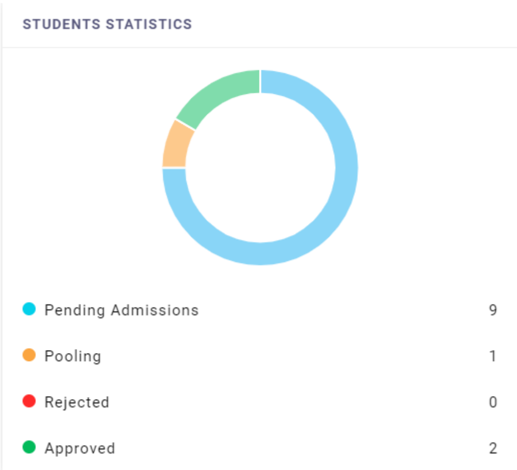


Figure 6.6 School statistics

Recently added students

ID	Student Name	School Name	Status	Date Added
1	Sadhik	ABC private school	Pooling	14/06/2020
2	Sathick	Al Najah Private School	Approved	19/06/2020
3	Test 4	Al Najah Private School	Rejected	20/06/2020
4	846	ABC private school	Approved	23/06/2020

Figure 6.7 Student status

As shown in Figure 6.8, the admission process of students follows a series of steps.

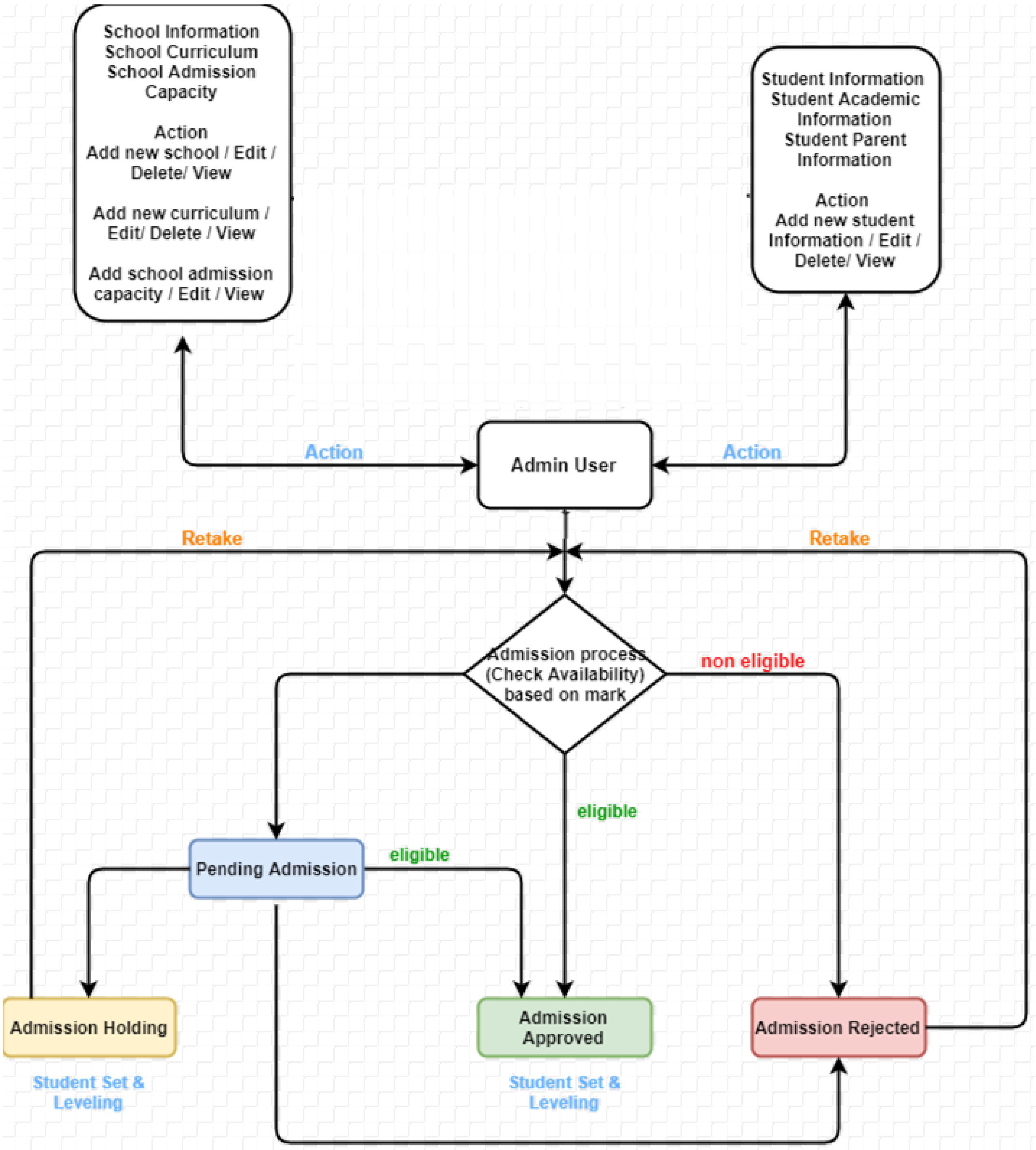


Figure 6.8 Admissions process implementation and simulation

6.2.3 Security and Privacy

The system is restricted to the school and teachers and protected with high security through a secure login page. The main reason is that any information or data related to a student or school record is sensitive and private. With a privacy-enhanced and security-enhanced student management web-based system, this kind of platform supports a low cost, time-effective, well-performed and secure application solution for schools. Therefore, the back-end objective is to improve admission process systems that work as a databank to save all the electronic student records information in a suitable method. Any school can retrieve, update, add, and delete their records and share the whole document if required by MOE.

However, accounts need to be highly secure; the password is vital to be protected. Instead of using traditional methods, which can be reversed quickly, the web-based system is designed to protect schools' and teachers' passwords using a salted password based on hashing technique [148]. Hash methods are developed with one-way tasks. It works by converting any quantity of data into a fixed length that is impossible to reverse if hackers attempt to obtain any meaningful information. Ideally, using hash processes is the optimal method to protect passwords, which can't convert a hash code back into its original string [149]. Therefore, there is a high possibility that hackers and malicious applications may attempt to utilise brute-force attacks. A "salting" function has been added to avoid that situation that can provide a random string known as salt to the password [149].

As mentioned earlier, the web-based system deals with sensitive student and school data in the central database accessed by authorised individuals who can use the system. Teachers and schools must have the correct username and password sent by the main administrator to access the system. Figure 6.9 demonstrates the log in page.

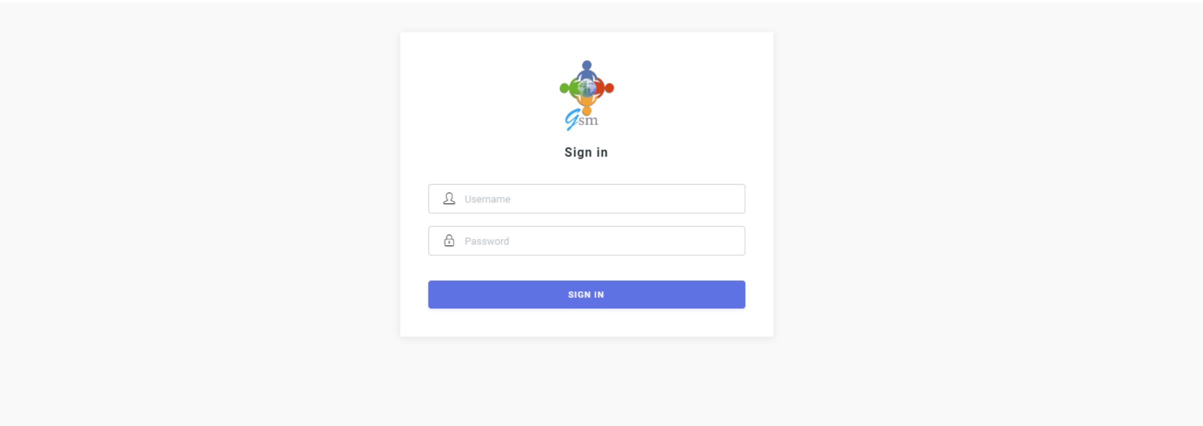


Figure 6.9 System log in page

6.3 Simulation and Evaluations

The admission process in schools is done using simple essential tools by the admission officer; therefore, to test the accuracy of the system and its efficiency, we simulated admissions for some students in a British school. The results shown in Table 6.1 compares the proposed year group calculated manually by the admission officer to the proposed year group generated by the system. The system-generated results align with the admission officer ones, which approves the system's accuracy. However, in some cases, there were discrepancies between the system and the admission officer's decisions, as illustrated in Table 6.2, which shows that three of the students (Year 1, 7, and 9) were assigned lower year than the system's results. System's results for the three special cases have been communicated to the school and the school's management has provided me the parents' undertakings which approve that parents agree on repeating the previous academic year for their children as they came from South Africa that ends the academic year by December. Undertaking letters and system's results for the special cases are attached in Appendix C.

Table 6.1 Baseline comparison table for admission model testing

Student ID	DOB	Admission Date	Previous Curriculum	Proposed Curriculum	Previous Year/Grade Group	Proposed Year/Grade Group by School Admission Officer	System Generated Results
846	09/05/13	SEP-20	British	British	Y2	Y3	**Y3**
847	08/10/13	SEP-20	British	British	Y1	Y2	**Y2**
848	02/10/13	SEP-20	British	British	Y2	Y3	**Y3**
849	30/07/13	SEP-20	British	British	Y2	Y3	**Y3**
850	15/02/13	SEP-20	British	British	Y2	Y3	**Y3**
851	17/05/13	SEP-20	MOE	British	Y2	Y3	**Y3**
852	02/09/13	SEP-20	British	British	Y2	Y3	**Y3**
853	19/07/13	SEP-20	MOE	British	Y2	Y3	**Y3**
854	20/10/13	SEP-20	British	British	Y1	Y2	**Y2**
855	23/01/13	SEP-20	British	British	Y2	Y3	**Y3**
856	03/11/13	SEP-20	British	British	Y1	Y2	**Y2**

Table 6.2 Comparison table for admission model testing (cases)

Student ID	DOB	Admission Date	Previous Curriculum	Proposed Curriculum	Previous Year/Grade Group	Proposed Year/Grade Group by School Admission Officer	System Generated Results
54681	15/05/13	Apr-19	South African	British	KG1	FS2	**Y1**
45456	13/12/05	Apr-19	South African	British	G8	Y8	**Y9**
53567	07/01/07	Apr-19	South African	British	G5	Y6	**Y7**

Whilst simulations were done on the accuracy of the system generated results, we timed every admission process to evaluate how efficient the system is compared to that manually done in the school. As shown in Figure 6.10 and Figure 6.11, the system was more efficient than the process currently used in schools concerning time to decide.

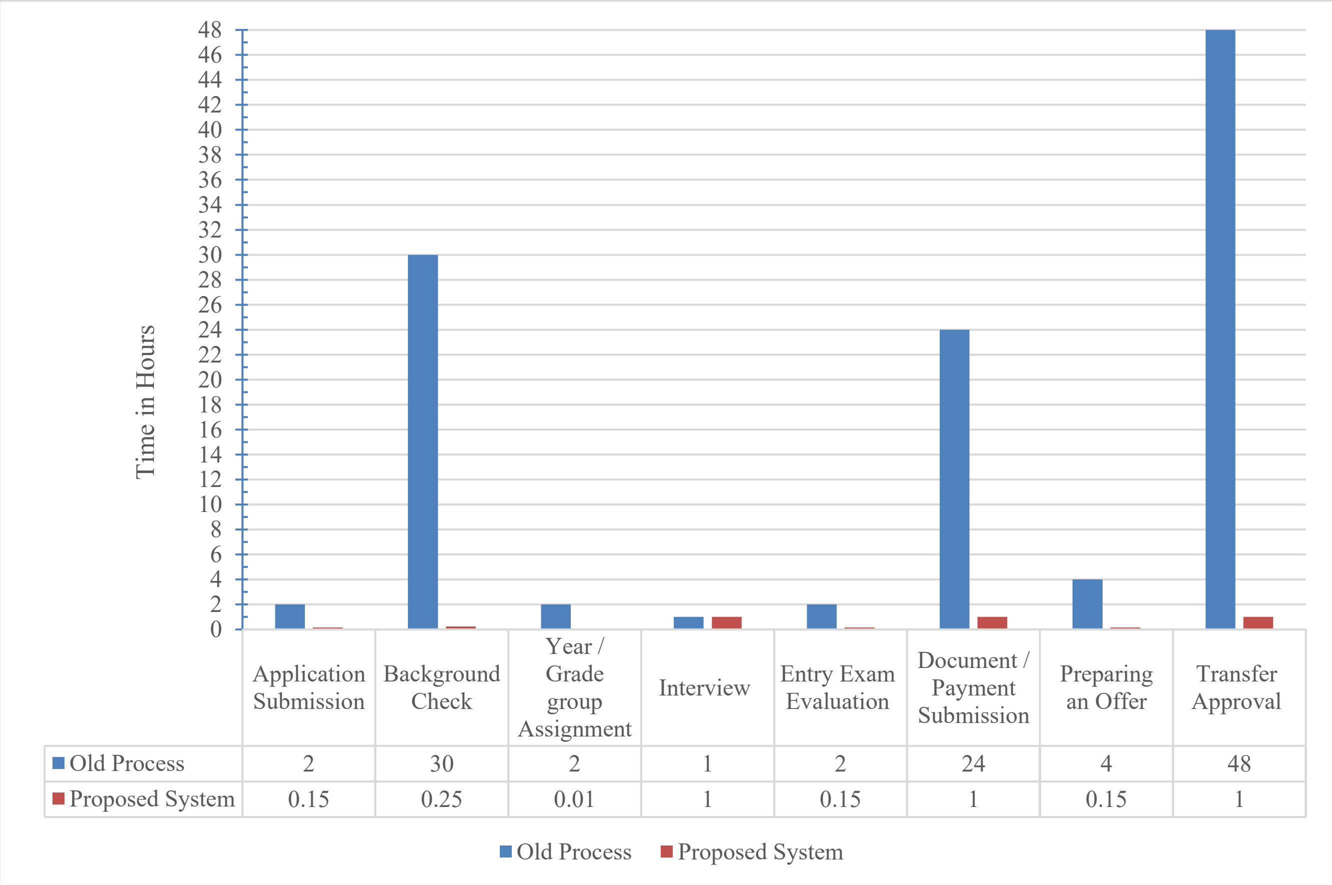


Figure 6.10 Time consumption of old admission process vs proposed system model

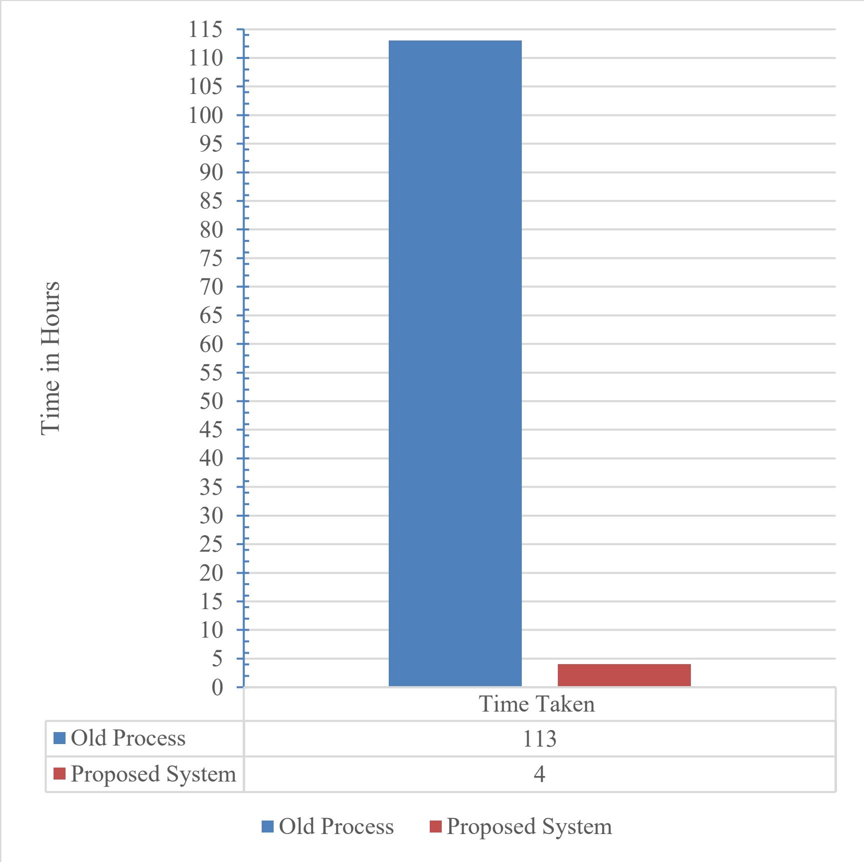


Figure 6.11 Total time comparison of old admission processes vs proposed system model

6.4 Chapter Summary

This chapter discussed the details of the student web-based system used to support the proposed research. A novel web-based system that manages students' data was presented when transferring between schools, providing them with their level according to their exam marks. Understanding the criteria followed by all schools in Abu Dhabi, UAE, has been challenging as each school has its procedures. However, based on gathered information received from schools and MOE, the initial target of managing student admissions was proposed in this system. This chapter reviewed the effectiveness of implementing such a system into schools and having one unified system to store student data.

Chapter 7 Conclusion and Future Work

7.1 Thesis Summary

This research recommends applying artificial intelligence technology to enhance the education quality and schools' admission in multicultural countries by predicting and automating the most suitable students' levels when transferring between different curriculums. This study has focused on three significant perspectives to improve the way of student levelling, differentiation, and the admission processes. First, the study used machine learning algorithms based on collected real students' datasets to predict the correct level of students. Second, the study supports parents during the admission process by saving their time and efforts and assessing the students by predicting their appropriate level. Finally, this research study has designed a user-friendly platform based on a web-based student management system to bring both perspectives together in one platform for schools and parents. It is a unified platform that holds students' data and management throughout their academic journey.

Implementing machine learning for the classification process could help education providers to predict the correct level of students more efficiently without the need for regular examinations as they can learn from data that has been previously collected. This research shows that the Artificial Neural Network method can generate efficient predictions of students' performance from the collected dataset. Various machine learning techniques for classification, including Artificial Neural Network and Random Forest, were used in this study alongside combined classifiers. Extensive simulation results indicated that the best average results had been given by the LEVNN method, which is one of the methods used for training the Neural Network to solve nonlinear least-squares problems.

International schools in Abu Dhabi are not following the same grading system. Therefore, the level of the students is affected or given incorrectly due to an inconsistent levelling system. Using the proposed machine learning model alongside the web-based system can help the levelling process effectively and efficiently. The suggested method addressed the issues stated in chapter 2 because there is no capable model to predict the level of the students transferring from different curricula. The proposed model resolves the problem of not having a unified student levelling system.

7.2 Achievements

7.3 Research Contributions

The contribution of the research can be assessed from various perspectives, including machine learning and web-based system combined with educational and IT domains. The study focuses on how machine learning can predict students' levels in the future. In addition, it has added further innovations in the field of machine learning models related to education. Other researchers can benefit from the research techniques and the dataset collected and manipulated for their studies.

UAE is a multicultural country with many expatriates relocating from Asia, Europe, and America. To meet expatriates needs, UAE has established many international private schools. However, since every country has a different curriculum, many challenges were faced by schools and MOE in allocating students to their correct year/group, keeping track of their academic performance relocating between schools, and assigning them to their proper level. Consequently, these data are essential to show student levelling faced by schools and MOE in different curriculums. Also, these data help highlight how students' levels can vary when they transfer between curriculums. The dataset comprises novel aspects of student grading in diverse educational cultures within multiple countries. Researchers and other education sectors can use this data to see the impact of having varied curriculums in a country. The dataset can be used by machine learning algorithms and pattern analysis methods to develop an intelligent framework applicable to multicultural educational systems. It can aid in a smooth transition "levelling", hereafter of students who relocate from a particular education curriculum to another and minimize the impact of switching on the students' educational performance.

The exam marks of students from their educational records were collected for each term for two consecutive academic years. The exam marks were collected for three terms of two academic years from Math, Science, and English significant courses. However, before admitting students into schools, the corresponding records stored in schools that include entry exam marks, nationality, and schooling system they came from, were also collected. Following the Ethical approval, after receiving a permission letter from the UAE Ministry of Education, British and American schools were contacted to arrange a meeting. The school principal granted access to the gatekeeper for data collection. The data from the British and American schools were collected in excel format, which was done entirely by the concierge. Dataset was updated periodically when needed.

The main objectives of this research have been achieved through reviewing previous work of different education curricula applied in different countries like the culture of the UAE. Also, by reviewing several works of literature on the applicability of applying Machine Learning in the education field to enhance the understanding of this area and choose the best model for the proposed research. On the other hand, I conducted an exploratory data analysis to select the relevant features that would assist in student levelling. And designed a Machine Learning based framework to analyse students' data and infer the ideal future level. Then I conducted a testing process the proposed tool and viability of using these technologies via several experiments and using reliable performance measurements. Finally, I developed a rule-based system that integrates the rules of admission and the ML predictions to automate proper decisions of student levelling.

This study selected different models to discover the best classifiers to generate the best accuracy and performance, as discussed in chapter 4. The standard performance measurements such as Sensitivity, Specificity, Precision, J1, F1 score, Accuracy, AUC, and ROC have been implemented to measure the performance of the classifiers. Various simulations are conducted on student levelling datasets to evaluate the models discussed in chapters 3 and 4. LEVNN model produced the best performance results amongst all the models applied in this research during the testing process. Overall, the results gained highlight the importance and potential of student data for student levelling.

7.4 Future Research Directions

With the experimental study's success, this research considers further work trends, including advancements to the proposed machine learning models (single and ensemble classifiers) and the web-based student management system. The study used RFC and ANN as single classifiers and combined NN Com, LEVNN Com, and NN and RF Com1 and Com2 with supervised learning. Future work can use the global optimisation algorithms to explore more comprehensively the space of possible machine learning architectures. It is noted that the current study has addressed only a limited set of architectures, which may not expose the full potential of the machine learning algorithms within the classification setting; this research suggested therefore that an algorithmic model search may be used to expand the scope and scale of this study. It is also noticed that the main limitation of the proposed models is computational performance. More algorithms can be used in supervised learning techniques such as SVM, K-nearest-neighbours, and Naïve Bays. Unsupervised learning of clustering analysis using K-means, Hidden Markov Models, Neural Networks, and Gaussian Mixture would improve students' levelling system research [150].

Collecting more records for students from different schools with other curricula and additional academic years would improve the results. As many curricula are applied in the UAE, such as Local, Indian, Pakistani, Australian, and Canadian curricula, it is recommended to include them for expanding the research. In addition, increasing features scope by adding different lifestyle attributes will be great for future studies. These features could include students' learning style, motivation and interest, family background, personality type, and information processing ability.

This research concentrates on students' levelling predictions when they transfer between curricula. Another direction for improvement of the study is to give more attention to grouping and differentiation systems in schools where students' periodic test marks are considered for predicting each student's coming level in a specific subject. Differentiation is a form of levelling that is hardly found in schools in the UAE. Every student currently shares the same work in class and at home without guides to identify weak and strong students in each topic. Students who are weak in certain subjects are given the same job as stronger ones, which creates a significant gap in their learning. Machine Learning can solve this problem by predicting the correct group using their historical marks of each course.

On the other hand, the web-based system to be amended in the future according to the future work of differentiation as mentioned above; this might include adding new features and rules to the system to cope with this process that match the differentiation requirements. Also adding more curricula than only the UK and the US ones will need an amendment to the rule-based system as well to align with this addition; this will include adding new rules according to these curricula.

REFERENCES

[1] M. E. Webb *et al.*, “Machine learning for human learners: opportunities, issues, tensions and threats,” *Educational Technology Research and Development*, vol. 69, no. 4, pp. 2109–2130, 2021, doi: 10.1007/s11423-020-09858-2.

[2] M. W. Lee, S. Y. Chen, K. Chrysostomou, and X. Liu, “Mining students’ behavior in web-based learning programs,” *Expert Syst Appl*, vol. 36, no. 2 PART 2, pp. 3459–3464, 2009, doi: 10.1016/j.eswa.2008.02.054.

[3] K. T. 8JB Profile Locations, Spray Hill, Hastings Road, Lamberhurst, “Managing the challenges of moving between school curricula,” *Relocate magazine*, 2022. https://www.relocatemagazine.com/articles/3522moving-between-international-curriculum-choices-managing-the-challenges (accessed Jul. 13, 2022).

[4] M. Khalaf *et al.*, “Applied Difference Techniques of Machine Learning Algorithm and Web-Based Management System for Sickle Cell Disease,” *Proceedings - 2015 International Conference on Developments in eSystems Engineering, DeSE 2015*, pp. 231–235, 2016, doi: 10.1109/DeSE.2015.18.

[5] J. Mao, “Why artificial neural networks - Mao - 1996.pdf,” 1996.

[6] R. L. Hale, “Cluster analysis in school psychology: An example,” *J Sch Psychol*, vol. 19, no. 1, pp. 51–56, 1981, doi: 10.1016/0022-4405(81)90007-8.

[7] M. A. B. Ahmad, “Mining Health Data for Breast Cancer Diagnosis Using Machine Learning,” *Austrailia university of Canberra*, no. December, 2013.

[8] M. Huber, C. Kurz, and R. Leidl, “Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning,” *BMC Med Inform Decis Mak*, vol. 19, no. 1, pp. 1–13, 2019, doi: 10.1186/s12911-018-0731-6.

[9] D. G. Chandra and D. Borah Malaya, “Role of cloud computing in education,” *2012 International Conference on Computing, Electronics and Electrical Technologies, ICCEET 2012*, pp. 832–836, 2012, doi: 10.1109/ICCEET.2012.6203884.

[10] N. Mac Parthaláin, “Rough Set Extensions for Feature Selection Rough Set Extensions for Feature Selection Neil S . Mac Parthal ´ Department of Computer Science,” no. September 2009, 2014.

[11] I. Conference and O. N. Electronics, “Reviews using Supervised Learning Algorithm,” *Ieee*, pp. 548–552, 2015.

[12] C. Masci, G. Johnes, and T. Agasisti, “Student and school performance across countries: A machine learning approach,” *European Journal of Operational Research*, vol. 269, no. 3. pp. 1072–1085, 2018. doi: 10.1016/j.ejor.2018.02.031.

[13] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, “Machine learning approaches to predict learning outcomes in Massive open online courses,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, pp. 713–720, 2017, doi: 10.1109/IJCNN.2017.7965922.

[14] T. C. Hsia, A. J. Shie, and L. C. Chen, “Course planning of extension education to meet market demand by using data mining techniques - an example of Chinkuo technology university in Taiwan,” *Expert Syst Appl*, vol. 34, no. 1, pp. 596–602, 2008, doi: 10.1016/j.eswa.2006.09.025.

[15] C. Masci, G. Johnes, and T. Agasisti, “Student and school performance across countries: A machine learning approach,” *European Journal of Operational Research*, vol. 269, no. 3. pp. 1072–1085, 2018. doi: 10.1016/j.ejor.2018.02.031.

[16] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, “Machine learning approaches to predict learning outcomes in Massive open online courses,” *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, pp. 713–720, 2017, doi: 10.1109/IJCNN.2017.7965922.

[17] T. C. Hsia, A. J. Shie, and L. C. Chen, “Course planning of extension education to meet market demand by using data mining techniques - an example of Chinkuo technology university in Taiwan,” *Expert Syst Appl*, vol. 34, no. 1, pp. 596–602, 2008, doi: 10.1016/j.eswa.2006.09.025.

[18] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, “Dropout prediction in e-learning courses through the combination of machine

learning techniques," *Comput Educ*, vol. 53, no. 3, pp. 950–965, 2009, doi: 10.1016/j.compedu.2009.05.010.

[19] R. R. Colditz, "An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms," *Remote Sens (Basel)*, vol. 7, no. 8, pp. 9655–9681, 2015, doi: 10.3390/rs70809655.

[20] J. Campbell, "Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study.," 2007.

[21] N. Campasano, R. Bergin, F. Approved, and O. M. B. No, "NAVAL POSTGRADUATE THESIS COMMUNITY PREPAREDNESS : by," no. March, 2010.

[22] A. Zakaryazad and E. Duman, "A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing," *Neurocomputing*, vol. 175, no. PartA, pp. 121–131, 2016, doi: 10.1016/j.neucom.2015.10.042.

[23] S. Ghareeb, A. Hussain, W. Khan, D. Al-Jumeily, T. Baker, and R. Al-Jumeily, "Dataset of student level prediction in UAE," *Data Brief*, vol. 35, p. 106908, 2021, doi: 10.1016/j.dib.2021.106908.

[24] S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades," *Artif Intell Rev*, vol. 37, no. 4, pp. 331–344, 2012, doi: 10.1007/s10462-011-9234-x.

[25] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[26] G. Zhang, "Support vector machines with Huffman tree architecture for multiclass classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3773 LNCS, no. August 2014, pp. 24–33, 2005, doi: 10.1007/11578079_3.

[27] M. Akour, H. al Sghaier, and O. al Qasem, “The effectiveness of using deep learning algorithms in predicting students achievements,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 388–394, 2020, doi: 10.11591/ijeecs.v19.i1.pp388-394.

[28] M. Akour, H. al Sghaier, and O. al Qasem, “The effectiveness of using deep learning algorithms in predicting students achievements,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 388–394, 2020, doi: 10.11591/ijeecs.v19.i1.pp388-394.

[29] Z. Karimi, M. Mansour Riahi Kashani, and A. Harounabadi, “Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods,” *Int J Comput Appl*, vol. 78, no. 4, pp. 21–27, 2013, doi: 10.5120/13478-1164.

[30] J. A. B. Fortes, “Sky Computing: When Multiple Clouds Become One,” pp. 4–4, 2010, doi: 10.1109/ccgrid.2010.136.

[31] M. Ramaswami and R. Bhaskaran, “A Study on Feature Selection Techniques in Educational Data Mining,” vol. 1, no. 1, pp. 7–11, 2009.

[32] S. Sivakumar, S. Venkataraman, and R. Selvaraj, “Predictive modeling of student dropout indicators in educational data mining using improved decision tree,” *Indian J Sci Technol*, vol. 9, no. 4, pp. 1–5, 2016, doi: 10.17485/ijst/2016/v9i4/87032.

[33] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. S. H. Rizvi, “A study of feature selection algorithms for predicting students academic performance,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 541–549, 2018, doi: 10.14569/IJACSA.2018.090569.

[34] J. R. Quinlan, “Induction of decision trees,” *Mach Learn*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/bf00116251.

[35] H. Ezzikouri, H. Liu, B. Hssina, A. Merbouha, and M. Erritali, “A comparative study of decision tree ID3 and C4.5 Int ernat ional Journal of Dat a Mining & Knowledge Management Process (IJDKP), Chowdhury… Induct ion of Classificat ion Rules by Gini-Index Based Rule Generat ion A comparative study of decision tree ID3 and C4.5.” [Online]. Available: www.ijacsa.thesai.org

[36] M. ROBNIK SIKONJA MarkoRobnik and friuni-ljsi IGOR KONONENKO IgorKononenko, “Theoretical and Empirical Analysis of ReliefF and RReliefF,” *Mach Learn*, vol. 53, pp. 23–69, 2003, [Online]. Available: http://lkm.fri.uni-lj.si/xaigor/slo/clanki/MLJ2003-FinalPaper.pdf

[37] Nagesh Singh Chauhan, “Decision Tree Algorithm, Explained.” [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[38] Y. Y. Song and Y. Lu, “Decision tree methods: applications for classification and prediction,” *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[39] J. R. Quinlan, *C4.5: Programs for Machine Learning*. 1993.

[40] J. R. Quinlan, “Induction of decision trees,” *Mach Learn*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/bf00116251.

[41] Y. Yuan and M. J. Shaw, “Induction of fuzzy decision trees,” *Fuzzy Sets Syst*, vol. 69, no. 2, pp. 125–139, 1995, doi: 10.1016/0165-0114(94)00229-Z.

[42] S. Sayad, “Comparing Different Classification Techniques in Credit Scoring.” [Online]. Available: www.ismartsoft.comiSmartsoft

[43] G. James, D. Witten, T. Hastie, and R. Tibshirani, “An Introduction to Statistical Learning with Applications in R,” 2013.

[44] Nagesh Singh Chauhan, “Decision Tree Algorithm, Explained.” [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[45] R. M. Mohana, C. K. K. Reddy, P. R. Anisha, and B. V. R. Murthy, “Random forest algorithms for the classification of tree-based ensemble,” *Mater Today Proc*, Feb. 2021, doi: 10.1016/j.matpr.2021.01.788.

[46] T. K. Ho, “Random Decision Forests Tin Kam Ho Perceptron training,” *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282, 1995.

[47] L. Breiman, “Randon Forests,” *Machinelearning202.Pbworks.Com*, pp. 1–35, 1999.

[48] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[49] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, “Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 38–53, 2015, doi: 10.1016/j.isprsjprs.2015.03.002.

[50] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[51] P. Fergus, A. Hussain, D. al Jumeily, D. S. Huang, and N. Bouguila, “Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms,” *Biomed Eng Online*, pp. 1–26, 2017, doi: 10.1186/s12938-017-0378-z.

[52] E. Goel and E. Abhilasha, “International Journal of Advanced Research in Random Forest : A Review,” vol. 7, no. 1, pp. 251–257, 2017, doi: 10.23956/ijarcsse/V7I1/01113.

[53] K. Spoon *et al.*, “Random Forests for Evaluating Pedagogy and Informing Personalized Learning,” vol. 8, no. 2, pp. 20–50, 2016.

[54] T. A. Cardona, “Predicting Student Degree Completion using Random Forest,” 2020.

[55] K. P. Moustris, I. K. Larissi, P. T. Nastos, K. V. Koukouletsos, and A. G. Paliatsos, “Development and application of artificial neural network modeling in forecasting PM10 levels in a Mediterranean city,” *Water Air Soil Pollut*, vol. 224, no. 8, 2013, doi: 10.1007/s11270-013-1634-x.

[56] Simplilearn, “What is Perceptron: A Beginners Guide for Perceptron,” Nov. 29, 2021. https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron?source=frs_category (accessed Dec. 08, 2021).

[57] G. Asadollahfardi, "Artificial Neural Network," vol. 3, no. 1, pp. 77–91, 2015, doi: 10.1007/978-3-662-44725-3_5.

[58] A. Bulsari, "Some analytical solutions to the general approximation problem for feedforward neural networks," *Neural Networks*, vol. 6, no. 7, pp. 991–996, 1993, doi: 10.1016/S0893-6080(09)80008-7.

[59] J. L. Balcázar, R. Gavaldà, and H. T. Siegelmann, "Computational power of neural networks: A characterization in terms of Kolmogorov complexity," *IEEE Trans Inf Theory*, vol. 43, no. 4, pp. 1175–1183, 1997, doi: 10.1109/18.605580.

[60] R. Setiono and W. K. Leow, "FERNN: An algorithm for fast extraction of rules from neural networks," *Applied Intelligence*, vol. 12, no. 1–2, pp. 15–25, 2000, doi: 10.1023/A:1008307919726.

[61] N. Kasabov *et al.*, "Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: Design methodology and selected applications," *Neural Networks*, vol. 78, pp. 1–14, 2016, doi: 10.1016/j.neunet.2015.09.011.

[62] N. Gupta, "Artificial Neural Network," *Inter national Conference on Recent Trends in Applied Sciences with Engineering Applications*, vol. 3, no. 1, pp. 24–29, 2013, doi: 10.1007/978-3-662-44725-3_5.

[63] R. P. Lippmann, "An Introduction ' to Computing with Neural Nets".

[64] J. A. Yacim and D. G. B. Boshoff, "Impact of artificial neural networks training algorithms on accurate prediction of property values," *Journal of Real Estate Research*, vol. 40, no. 3, pp. 375–418, 2018, doi: 10.1080/10835547.2018.12091505.

[65] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network."

[66] S. Haykin, N. Networks, and A. C. Foundation, "BOOK REVIEW Exploring the Diversity of Artificial Neural Network Architectures," vol. 292, pp. 287–292, 1997.

[67] Y. Nojima and B. Gabrys, "Neural networks a comprehensive foundation".

[68] M. Shahin, M. B. Jaksa, and H. R. Maier, "State of the Art of Artificial Neural Networks in Geotechnical Engineering," no. April 2015, 2008.

[69] R. P. Lippmann, “An Introduction ’ to Computing with Neural Nets”.

[70] H. Al-Haj Ibrahim, “Fouling of heat exchangers,” *MATLAB – A Fundamental Tool for Scientific Computing and Engineering Applications* , vol. 3, no. 1, pp. 57–96, May 2012, doi: 10.1016/S0017-9310(96)00261-X.

[71] Y. C. Du and A. Stephanus, “Levenberg-marquardt neural network algorithm for degree of arteriovenous fistula stenosis classification using a dual optical photoplethysmography sensor,” *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072322.

[72] Z. Cömert and A. F. Kocamaz, “A study of artificial neural network training algorithms for classification of cardiotocography signals,” *JOURNAL OF SCIENCE AND TECHNOLOGY*, vol. 7, no. 2, pp. 93–103, 2017, [Online]. Available: www.dergipark.ulakbim.gov.tr/beuscitech/

[73] F. M. Alkoot and J. Kittler, “Experimental evaluation of expert fusion strategies,” vol. 20, 1999.

[74] R. Polikar, “Polikar, R.: Ensemble based systems in decision making. IEEE Circuit Syst. Mag. 6, 21-45,” no. October 2006, pp. 21–45, 2015, doi: 10.1109/MCAS.2006.1688199.

[75] R. E. Schapire, “The Strength of Weak Learnability,” vol. 227, pp. 197–227, 1990.

[76] L. E. O. Bbeiman, “Bagging Predictors,” vol. 140, pp. 123–140, 1996.

[77] K. C. Lee and H. Cho, “Performance of Ensemble Classifier for Location Prediction Task : Emphasis on Markov Blanket Perspective Performance of Ensemble Classifier for Location Prediction Task : Emphasis on Markov Blanket Perspective,” no. June, 2014.

[78] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. de Sousa, “Ensemble approaches for regression: A survey,” *ACM Computing Surveys*, vol. 45, no. 1. Nov. 2012. doi: 10.1145/2379776.2379786.

[79] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.

[80] J. Davis and M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves,” 2006.

[81] S. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “unc orre cted,” no. November, 2006, doi: 10.1007/s10462-007-9052-3.

[82] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, and J. Mononen, “Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle,” *Behavioural Processes*, vol. 148, pp. 56–62, Mar. 2018, doi: 10.1016/j.beproc.2018.01.004.

[83] K. Nighania, “Various ways to evaluate a machine learning model’s performance,” Dec. 2018. [Online]. Available: https://towardsdatascience.com/various-ways-to-evaluate-a-machine-lear...

[84] J. Johnson, “Precision, Recall & Confusion Matrices in Machine Learning,” Jul. 2020.

[85] ADEK, “Main page,” 2019. https://www.adek.gov.ae/ (accessed Sep. 17, 2019).

[86] J. Huang and C. X. Ling, “Using AUC and accuracy in evaluating learning algorithms,” *IEEE Trans Knowl Data Eng*, vol. 17, no. 3, pp. 299–310, Mar. 2005, doi: 10.1109/TKDE.2005.50.

[87] A. Ben-David and E. Frank, “Accuracy of machine learning models versus ‘hand crafted’ expert systems - A credit scoring case study,” *Expert Syst Appl*, vol. 36, no. 3 PART 1, pp. 5264–5271, 2009, doi: 10.1016/j.eswa.2008.06.071.

[88] R. Lotto, “Postgraduate students’ perception of research ethics training: a qualitative study,” 2018. [Online]. Available: http://openjournals.ljmu.ac.uk/iip

[89] “governanceoffacultyrecspaperupdatednovember2017”.

[90] D. H. Christiansen, J. D. Hosking, A. L. Dannenberg, and O. D. Williams, "Computer-Assisted Data Collection in Multicenter Epidemiologic Research The Atherosclerosis Risk in Communities Study*," 1990.

[91] R. Clarke, "Big data, big risks," *Information Systems Journal*, vol. 26, no. 1, pp. 77–90, Jan. 2016, doi: 10.1111/isj.12088.

[92] A. J. Hussain, P. Fergus, H. Al-Askar, D. Al-Jumeily, and F. Jager, "LJMU Research Online Dynamic Neural Network Architecture Inspired by The Immune Algorithm to Predict Preterm Deliveries in Pregnant Women", [Online]. Available: http://researchonline.ljmu.ac.uk/

[93] D. Kučak, V. Juričić, and G. Đambić, "Machine learning in education - A survey of current research trends," in *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, 2018, vol. 29, no. 1, pp. 0406–0410. doi: 10.2507/29th.daaam.proceedings.059.

[94] D. B. Lomet, L. Gravano, A. Levy, and G. Weikum, "Editorial Board Editor-in-Chief Associate Editors." [Online]. Available: http://list.research.microsoft.com/scripts/lyris.pl?enter=debull.

[95] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, "Video scene analysis: an overview and challenges on deep learning algorithms," *Multimed Tools Appl*, vol. 77, no. 16, pp. 20415–20453, Aug. 2018, doi: 10.1007/s11042-017-5438-7.

[96] Q. Guo and M. Zhang, "Implement web learning environment based on data mining," *Knowl Based Syst*, vol. 22, no. 6, pp. 439–442, 2009, doi: 10.1016/j.knosys.2009.06.001.

[97] M. Al-khafajiy, S. Ghareeb, R. Al-jumeily, and R. Almurshedi, "A Holistic Study on Emerging IoT Networking Paradigms," 2019.

[98] E. Parra, C. Dimou, J. Llorens, V. Moreno, and A. Fraga, "A methodology for the classification of quality of requirements using machine learning techniques," in *Information and Software Technology*, Nov. 2015, vol. 67, pp. 180–195. doi: 10.1016/j.infsof.2015.07.006.

[99] R. Mariescu-Istodor and I. Jormanainen, “Machine Learning for High School Students,” vol. 19, 2019, doi: 10.1145/3364510.

[100] “EXPLORATORY DATA ANALYSIS.”

[101] M. C. Cieslak, A. M. Castelfranco, V. Roncalli, P. H. Lenz, and D. K. Hartline, “t-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis,” *Mar Genomics*, vol. 51, Jun. 2020, doi: 10.1016/j.margen.2019.100723.

[102] D. K. Agrafiotis, “Stochastic Proximity Embedding,” 2003.

[103] G. C. Linderman, M. Rachh, J. G. Hoskins, S. Steinerberger, and Y. Kluger, “Efficient Algorithms for t-distributed Stochastic Neighborhood Embedding,” Dec. 2017, doi: 10.1038/s41592-018-0308-4.

[104] J. Xia *et al.*, “LDSScanner: Exploratory Analysis of Low-Dimensional Structures in High-Dimensional Datasets,” *IEEE Trans Vis Comput Graph*, vol. 24, no. 1, pp. 236–245, Jan. 2018, doi: 10.1109/TVCG.2017.2744098.

[105] A. Smoli, B. Walczak, and J. W. Einax, “Exploratory analysis of data sets with missing elements and outliers.” [Online]. Available: www.elsevier.com/locate/chemosphere

[106] D. K. Agrafiotis, H. Xu, F. Zhu, D. Bandyopadhyay, and P. Liu, “Stochastic proximity embedding: Methods and applications,” *Mol Inform*, vol. 29, no. 11, pp. 758–770, Nov. 2010, doi: 10.1002/minf.201000134.

[107] D. N. Rassokhin and D. K. Agrafiotis, “A modified update rule for stochastic proximity embedding,” *J Mol Graph Model*, vol. 22, no. 2, pp. 133–140, 2003, doi: 10.1016/S1093-3263(03)00155-4.

[108] G. Aloor and L. Jacob, “Distributed wireless sensor network localization using stochastic proximity embedding,” *Comput Commun*, vol. 33, no. 6, pp. 745–755, Apr. 2010, doi: 10.1016/j.comcom.2009.11.022.

[109] S. Agarwal, "Data mining: Data mining concepts and techniques," in *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, Oct. 2014, pp. 203–207. doi: 10.1109/ICMIRA.2013.45.

[110] J. van den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst, "Data cleaning: Detecting, diagnosing, and editing data abnormalities," *PLoS Medicine*, vol. 2, no. 10. pp. 0966–0970, 2005. doi: 10.1371/journal.pmed.0020267.

[111] N. R. Prasad, S. Almanza-Garcia, and T. T. Lu, "Anomaly detection," *Computers, Materials and Continua*, vol. 14, no. 1, pp. 1–22, 2009, doi: 10.1145/1541880.1541882.

[112] by Ahmed Jasim Mohammed Kaky, "INTELLIGENT SYSTEMS APPROACH FOR CLASSIFICATION AND MANAGEMENT OF PATIENTS WITH HEADACHE," 2017.

[113] Z. Zhang, "Missing data imputation: Focusing on single imputation," *Ann Transl Med*, vol. 4, no. 1, 2016, doi: 10.3978/j.issn.2305-5839.2015.12.38.

[114] Z. Zhang, "Missing data imputation: Focusing on single imputation," *Ann Transl Med*, vol. 4, no. 1, Jan. 2016, doi: 10.3978/j.issn.2305-5839.2015.12.38.

[115] L. Moyé, "Statistical methods for cardiovascular researchers," *Circ Res*, vol. 118, no. 3, pp. 439–453, 2016, doi: 10.1161/CIRCRESAHA.115.306305.

[116] H. Kang, "The prevention and handling of the missing data," *Korean J Anesthesiol*, vol. 64, no. 5, pp. 402–406, 2013, doi: 10.4097/kjae.2013.64.5.402.

[117] Y. Li and A. Ngom, "Data integration in machine learning," in *Proceedings - 2015 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2015*, Dec. 2015, pp. 1665–1671. doi: 10.1109/BIBM.2015.7359925.

[118] P. Ziegler and K. R. Dittrich, "Data Integration-Problems, Approaches, and Perspectives."

[119] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naïve Bayes a Good Classifier for Document Classification?," 2011. [Online]. Available: https://www.researchgate.net/publication/266463703

[120] T. R. N and R. Gupta, “An efficient feature subset selection approach for machine learning,” *Multimed Tools Appl*, vol. 80, no. 8, pp. 12737–12830, Mar. 2021, doi: 10.1007/s11042-020-10011-7.

[121] T. Dietterich, “Overfitting and Undercomputing in Machine Learning.”

[122] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” no. April, 1999.

[123] H. Vafaie and K. De Jong, “Genetic algorithms as a tool for feature selection in machine learning,” *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, vol. 1992-Novem, no. January 2013, pp. 200–203, 1992, doi: 10.1109/TAI.1992.246402.

[124] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” no. April, 1999.

[125] M. Ramaswami and R. Bhaskaran, “A Study on Feature Selection Techniques in Educational Data Mining,” vol. 1, no. 1, pp. 7–11, 2009, [Online]. Available: http://arxiv.org/abs/0912.3924

[126] A. Jović, K. Brkić, and N. Bogunović, “A review of feature selection methods with applications,” in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, Jul. 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.

[127] S. Luo, “Data Mining of Many-Attribute Data: Investigating the Interaction between Feature Selection Strategy and Statistical Features of Datasets.”

[128] M. Q. Tran, M. Elsisi, K. Mahmoud, M. K. Liu, M. Lehtonen, and M. M. F. Darwish, “Experimental Setup for Online Fault Diagnosis of Induction Machines via Promising IoT and Machine Learning: Towards Industry 4.0 Empowerment,” *IEEE Access*, vol. 9, pp. 115429–115441, 2021, doi: 10.1109/ACCESS.2021.3105297.

[129] H. M. El-bakry, “An efficient algorithm for pattern detection using combined classifiers and data fusion,” *Information Fusion*, vol. 11, no. 2, pp. 133–148, 2010, doi: 10.1016/j.inffus.2009.06.001.

[130] H. M. El-bakry, “An efficient algorithm for pattern detection using combined classifiers and data fusion,” *Information Fusion*, vol. 11, no. 2, pp. 133–148, 2010, doi: 10.1016/j.inffus.2009.06.001.

[131] A. Jain and S. Solanki, “An Efficient Approach for Multiclass Student Performance Prediction based upon Machine Learning,” no. Icces, pp. 1457–1462, 2019.

[132] J. Aguirre-Gutiérrez, A. C. Seijmonsbergen, and J. F. Duivenvoorden, “Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico,” *Applied Geography*, vol. 34, pp. 29–37, 2012, doi: 10.1016/j.apgeog.2011.10.010.

[133] P. A. Flach PETERFLACH, “An Analysis of Rule Evaluation Metrics Johannes F ¨ urnkranz.”

[134] G. S. Handelman *et al.*, “Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods,” *American Journal of Roentgenology*, vol. 212, no. 1. American Roentgen Ray Society, pp. 38–43, Jan. 01, 2019. doi: 10.2214/AJR.18.20224.

[135] J.-O. Palacio-Niño and F. Berzal, “Evaluation Metrics for Unsupervised Learning Algorithms,” May 2019, [Online]. Available: http://arxiv.org/abs/1905.05667

[136] B. U. Ca and Y. G. Fr, “No Unbiased Estimator of the Variance of K-Fold Cross-Validation Yoshua Bengio Yves Grandvalet,” 2004.

[137] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, “Evaluating the quality of machine learning explanations: A survey on methods and metrics,” *Electronics (Switzerland)*, vol. 10, no. 5. MDPI AG, pp. 1–19, Mar. 01, 2021. doi: 10.3390/electronics10050593.

[138] I. H. Sarker, A. S. M. Kayes, and P. Watters, “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage,” *J Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0219-y.

[139] C. Ordonez, “Association rule discovery with the train and test approach for heart disease prediction,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 2, pp. 334–343, Apr. 2006, doi: 10.1109/TITB.2006.864475.

[140] L. Breiman, “Randon Forests,” *Machinelearning202.Pbworks.Com*, pp. 1–35, 1999, [Online]. Available: http://machinelearning202.pbworks.com/w/file/fetch/60606349/breiman_randomforests.pdf

[141] A. Belayneh, J. Adamowski, B. Khalil, and J. Quilty, “Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction,” *Atmos Res*, vol. 172–173, pp. 37–47, May 2016, doi: 10.1016/j.atmosres.2015.12.017.

[142] B. C. Ko, H. H. Kim, and J. Y. Nam, “Classification of Potential Water Bodies Using Landsat 8 OLI and a Combination of Two Boosted Random Forest Classifiers,” pp. 13763–13777, 2015, doi: 10.3390/s150613763.

[143] M. F. Sikder, M. J. Uddin, and S. Halder, “Predicting students yearly performance using neural network: A case study of BSMRSTU,” in *2016 5th International Conference on Informatics, Electronics and Vision, ICIEV 2016*, Nov. 2016, pp. 524–529. doi: 10.1109/ICIEV.2016.7760058.

[144] T. A. Walsh, P. McMinn, and G. M. Kapfhammer, “Automatic detection of potential layout faults following changes to responsive web pages,” in *Proceedings - 2015 30th IEEE/ACM International Conference on Automated Software Engineering, ASE 2015*, Jan. 2016, pp. 709–714. doi: 10.1109/ASE.2015.31.

[145] V. Novák and S. Lehmke, “Logical structure of fuzzy IF-THEN rules,” *Fuzzy Sets Syst*, vol. 157, no. 15, pp. 2003–2029, Aug. 2006, doi: 10.1016/j.fss.2006.02.011.

[146] L. Wang *et al.*, “Cloud computing: A perspective study,” in *New Generation Computing*, Apr. 2010, vol. 28, no. 2, pp. 137–146. doi: 10.1007/s00354-008-0081-5.

[147] S. Bhardwaj and L. Jain, “An Approach for Investigating Perspective of Cloud Software-as-a-Service (SaaS),” 2010.

[148] P. Sriramya and R. A. Karthika, “PROVIDING PASSWORD SECURITY BY SALTED PASSWORD HASHING USING BCRYPT ALGORITHM,” vol. 10, no. 13, 2015, [Online]. Available: www.arpnjournals.com

[149] P. Sriramya and R. A. Karthika, “PROVIDING PASSWORD SECURITY BY SALTED PASSWORD HASHING USING BCRYPT ALGORITHM,” vol. 10, no. 13, 2015, [Online]. Available: www.arpnjournals.com

[150] S. Zanero and G. Serazzi, “Unsupervised learning algorithms for intrusion detection,” *NOMS 2008 - IEEE/IFIP Network Operations and Management Symposium: Pervasive Management for Ubiquitous Networks and Services*, pp. 1043–1048, 2008, doi: 10.1109/NOMS.2008.4575276.

Appendix A: Letter of Permission

دائـــرة التعليم والمعرفة
DEPARTMENT OF EDUCATION
AND KNOWLEDGE

Date:08/08/2019

التاريخ : 08/08/2019

To:Private Schools Principals

السادة / مديري المدارس الخاصة المحترمين

Subject : Letter Of Permission

الموضوع : تسهيل مهمة باحثين

Dear Principals,

تحية طيبة وبعد،

The Department of Education and Knowlegde would like to express its gratitude for your generous efforts and sincere cooperation in serving our dear researchers.

يطيب لدائرة التعليم و المعرفة ان تتوجه لكم بخالص الشكر والتقدير لجهودكم الكريمة و التعاون الصادق لخدمة ابنائنا الباحثين

You are kindly requested to allow the researcher /**Shatha Ghareeb, to** complete his research on:

و نود اعلامكم بموافقة دائرة التعليم و المعرفة علي موضوع الدراسة التي سيجريها الباحث / **Shatha Ghareeb** بعنوان:

A Distributed Machine Learning Based Framework for Students' Educational Levelling in Multicultural Countries: UAE as a case study

A Distributed Machine Learning Based Framework for Students' Educational Levelling in Multicultural Countries: UAE as a case study

Please indicate your approval of this permission by facilitating her meetings with the sample groups at your resoected schools.

لذا يرجي التكرم بتسهيل مهام الباحث و مساعدتة علي إجراء الدراسة المشار إليها

For Futher information : please
Helmy Seada on 02/6150140

للاستفسار : يرجي الاتصال **بالسيد / حلمي**
هاتف 02/6150140

Thank you four ur cooperation.
Sincerely yours,

شاكرين لكم حسن تعاونكم
وتفضلوا بقبول خالص الاحترام و التقدير

أ.د. مسعود عبد الله بدري
مدير وحدة البحوث والتخطيط وقياس الأداء

ص ب 36005 أبوظبي، الإمارات العربية المتحدة PO Box 36005 Abu Dhabi, United Arab Emirates
T +971 (0)2 615 0000 F +971 (0)2 615 0602 E info@adec.ac.ae

Appendix B: Some Matlab Codes

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Main Loader
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Get current path
pathtohere = [fileparts(mfilename('fullpath')),'\'];

% Config
CONFIG.clearexistingfigs = true;
CONFIG.outputpath = [pathtohere,'..\outputs'];
%CONFIG.outputpath = 'IEEEproject/outputs';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
CONFIG.inputfile = 'src.mat';
CONFIG.inputvar = 'src';
CONFIG.loadtargindex = 16;
CONFIG.loadfeatsindex = [1:8,9:15];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

CONFIG.numClassBins = 3;
CONFIG.seedrng = true;
%CONFIG.seedrngval = 13;
% 130,133, 135 137(ok test, imb train) 144, 151 chosen.
CONFIG.seedrngval = 151;
CONFIG.oversample = false;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Matlab display environment cleanup options

if CONFIG.clearexistingfigs == true
   close all;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Optionally Seed the Random Number Generator
```

```
if(CONFIG.seedrng)
    rng('default');
    rng(CONFIG.seedrngval);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Load the Data:

lo = MatLoader({...
            MatLoader.RULE_READ_P,CONFIG.inputfile,...
                CONFIG.inputvar,CONFIG.loadfeatsindex,...
            MatLoader.RULE_READ_T,CONFIG.inputfile,...
                CONFIG.inputvar,CONFIG.loadtargindex...
            });

% Preprocess the targets to yield a classification problem
DS_inputs = lo.Patterns;
[ORIG_targets,step] = map2binIdxs(lo.Targets,CONFIG.numClassBins); %
convert to classes
nClassGen = length(unique(ORIG_targets));
display(['Classes Generated: ',num2str(nClassGen),...
    ' (bins: ',num2str(CONFIG.numClassBins),')'])
display(['Class Labels: ',mat2str(unique(ORIG_targets))])
display(['Original Response Value range: ',...
    num2str(min(lo.Targets)),':', num2str(max(lo.Targets))])
display(['Class Discretization increment: ',num2str(step)])



%hist(ORIG_targets,unique(ORIG_targets))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Surrogate Data (this will be made into a modular feature in future dev)
% -> for now, comment and uncomment as required.
%[npatterns,nfeats] = size(DS_inputs);
%DS_inputs = randn(npatterns,nfeats);

%size(DS_inputs)
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Oversampling via SMOTE

if(CONFIG.oversample == true)
    [DS_inputs, ORIG_targets] = SMOTE(DS_inputs, ORIG_targets);
end
N_examples = length(ORIG_targets);

% Shuffle the order of the examples
shuffledInxs = randperm(N_examples);
ORIG_targets = ORIG_targets(shuffledInxs,:);
DS_inputs = DS_inputs(shuffledInxs,:);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Run Simulations

% Write to input buffer
INPUT_PATTERNS = DS_inputs;
INPUT_TARGETS = ORIG_targets;

%explore_data(INPUT_PATTERNS,INPUT_TARGETS);

% Call simulations script
runSimulations;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

prt = prtime();
prtime(inf);

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC50/1',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],50,1)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC100/2',...
```

```
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],100,2)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC200/3',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],200,3)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC400/4',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],400,4)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC500/5',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],500,5)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','Random Forest Classifier',...
    'shortname','RFC1000/6',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)C*randomforestc([],750,6)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','NN Com',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)
C*([lmnc([],2),bpxnc,lmnc([],5),lmnc([],10),lmnc([],20)]*lmnc)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
```

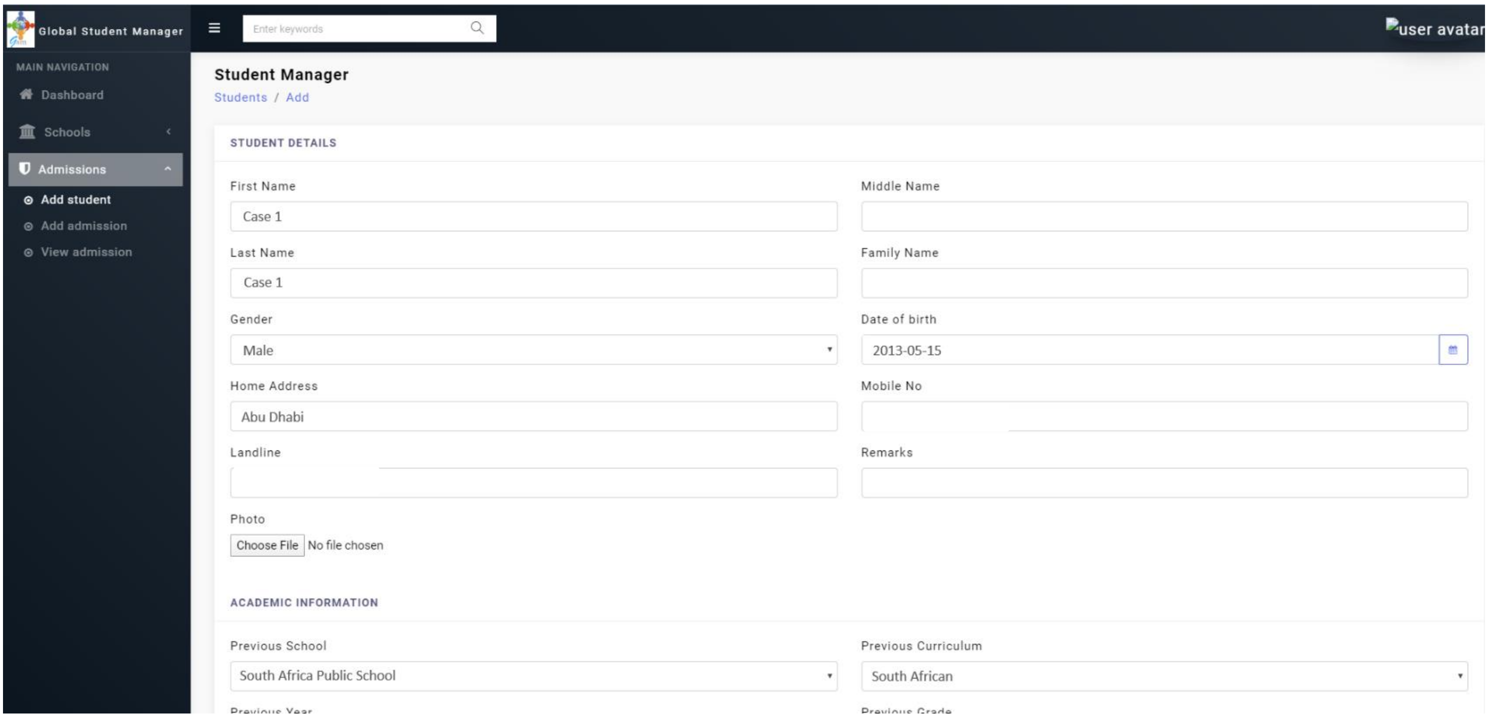
```
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','LEVNN Com',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)
C*([lmnc([],10),lmnc([],20),lmnc([],30),lmnc([],50)]*lmnc)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','NN and RF Com1',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C) C*([lmnc([],10),lmnc([],20),bpxnc,
lmnc([],30),lmnc([],50)]*randomforestc)...
    )];

% PR Models Set 1

prModelResultsArr = [prModelResultsArr,struct(...
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','NN Com',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)
C*([lmnc([],2),bpxnc,lmnc([],5),lmnc([],10),lmnc([],20)]*lmnc)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','LEVNN Com',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)
C*([lmnc([],10),lmnc([],20),lmnc([],30),lmnc([],50)]*lmnc)...
    )];

prModelResultsArr = [prModelResultsArr,struct(...
    'name','LEVNN Combined (LMNN combiner)',...
    'shortname','NN and RF Com1',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C) C*([lmnc([],10),lmnc([],20),bpxnc,
lmnc([],30),lmnc([],50)]*randomforestc)...
    )];


prModelResultsArr = [prModelResultsArr,struct(...
```

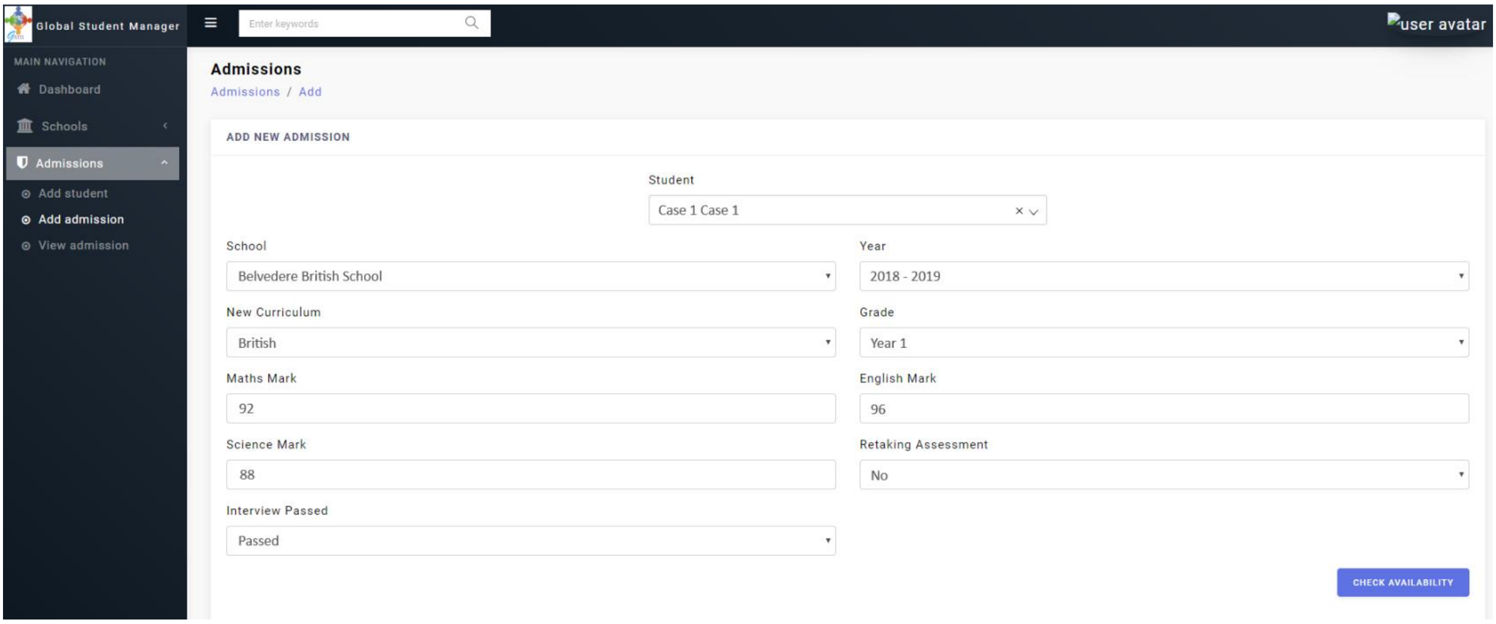
```
    'name','Neural Newtwork and Random Forest',...
    'shortname','NN and RF Com2',...
    'threshold',CONFIG.classthreshold,...
    'fcn',@(C)
C*([lmnc([],2),lmnc([],5),bpxnc,lmnc([],10),lmnc([],20)]*randomforestc)...
    )];
% Explore Data

% -----------------------------------------------------------------------
% Config

CONFIG.datafile = 'src.mat';
CONFIG.varname = 'src';
CONFIG.colsel = [1:4,8:15];
CONFIG.nbounds = [1,2];

% -----------------------------------------------------------------------
% Load Data

opt.colsel = CONFIG.colsel;
lo = QloadMat(CONFIG.datafile,CONFIG.varname,opt);

% -----------------------------------------------------------------------
% Unpack and Normalise just the patterns

p = lo.Patterns;
p = normaliseMatVars(p,CONFIG.nbounds(1),CONFIG.nbounds(2),0);
t = lo.Targets;
tc = map2binIdxs(t,CONFIG.numClassBins); % convert to classes

% -----------------------------------------------------------------------
% Characterise and Explore Data

% Feature Exploration
plot_ptscatters(p,t)

% Target Histogram
figure()
hist(t,20)
```

```
% PCA PLOT
figure()
[s,c] = pca(p);
%scatter(s(:,1),s(:,2));
gscatter(s(:,1),s(:,2),tc,'rgbcmrk','+o*.xsd')
title('PCA')

[dp, mapping] = compute_mapping(p, 'SPE', 2);
figure()
gscatter(dp(:,1),dp(:,2),tc,'rgbcmrk','+o*.xsd')
title('SPE plot for Sickle Cell Data')

[dp, mapping] = compute_mapping(p, 'tSNE', 2);
figure()
gscatter(dp(:,1),dp(:,2),tc,'rgbcmrk','+o*.xsd')
title('tSNE plot for Sickle Cell Data')
```

Appendix C: Special Cases

<u>Student ID: 54681</u>





BELVEDERE
British School

Parent Undertaking

I hereby confirm that I am in full agreement with School Name...Belvedere...British School
Which follows the ..British.. curriculum year system to repeat year....F.S 2 for my child for the academic year 2018/2019

Student Name: ...
Date of birth: .....
eSIS Student ID:......N A

I understand that the normal grade progression would place my child in Year ..F.S.2.for the academic year 2018/2019 , however I am requesting that my child repeat Year

تعهد ولي الأمر

أتعهد بموجب أدناه أنني أوافق كليا مع مدرسة.. التي تطبق المنهاج .... نظام ...... بإعادة السنة .... لابني/ابنتي للعام الدراسي 2019/2018

إسم الطالب:......
تاريخ الميلاد: ....
رقم الطالب على أيسس: ......

أتفهم أن التسلسل الدراسي الصحيح سيضع طفلي في السنة.......... للعام الدراسي 2019/2018 ومع ذلك أطلب أن يعيد أبني السنة .... نظرا للأسباب أدناه:

-1
-2

.... Because of -3
the below Reason:
1- We emigrated from South Africa in February
2- and they completed the school year in December
3-

I confirm that I will not be in a position to approach the Department of Education and Knowledge or the school to promote my child to a higher grade/year in the future.

كما أتعهد أنني لن أطالب دائرة التعليم والمعرفة أو المدرسة بترفيع ابني / ابنتي مستقبلا إلى صف أعلى.

I also understand that to meet the requirements of the Ministry of Education successfully in the British curriculum it will need to be completed until year 13.

أتفهم أنه لتحقيق متطلبات معادلة وزارة التربية والتعليم بنجاح في المنهاج البريطاني فإنه سيحتاج إلى إكمال حتى السنة 13.

Parent Name:
Signature:
Mobile Number:
Date:

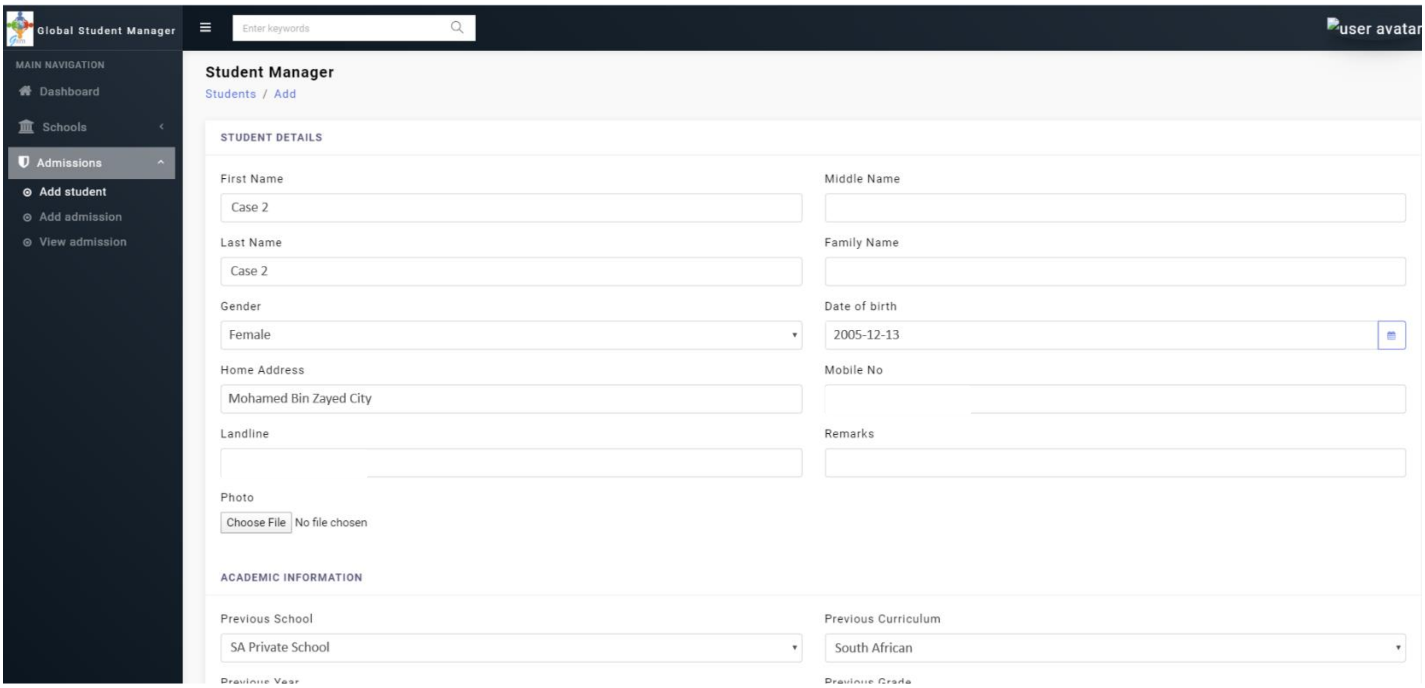
اسم ولي الأمر:
التوقيع:
رقم الموبايل:
التاريخ:

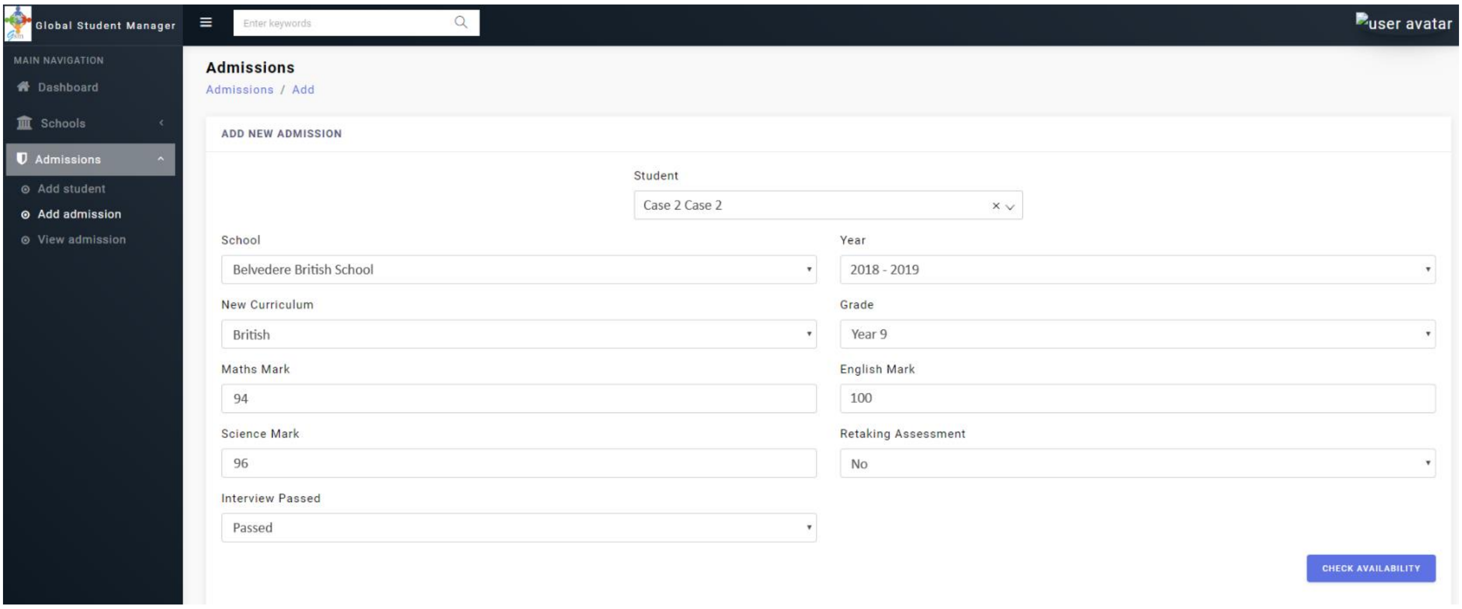
School Principal Decision

قرار مدير المدرسة

BELVEDERE British School
Abu Dhabi

<u>Student ID: 45456</u>





BELVEDERE
British School

Parent Undertaking

تعهد ولي الأمر

I hereby confirm that I am in full agreement with School Name...Belvedere... British School. Which follows the ...British curriculum year system to repeat year...8.. for my child for the academic year 2018/2019

أتعهد بموجب أدناه أنني أوافق كليا مع
مدرسة....................................... التي تطبق المنهاج
.... نظام ...... بإعادة السنة .... لابني/ابنتي للعام
الدراسي 2019/2018

Student Name: .
Date of birth: ...
eSIS Student ID:........

إسم الطالب:.......
تاريخ الميلاد: ....
رقم الطالب على أيسس: ......

I understand that the normal grade progression would place my child in Year ...8...for the academic year 2018/2019 , however I am requesting that my child repeat Year

أتفهم أن التسلسل الدراسي الصحيح
سيضع طفلي في السنة.......... للعام
الدراسي 2019/2018 ومع ذلك أطلب أن
يعيد أبني السنة .... نظرا للأسباب
أدناه:

-1
-2

...8..... Because of the below Reason: -3

1- we Emigrated from South Africa in February

2- and they finished the school year in December

3-

I confirm that I will not be in a position to approach the Department of Education and Knowledge or the school to promote my child to a higher grade/year in the future.

كما أتعهد أنني لن أطالب دائرة التعليم والمعرفة أو المدرسة بترفيع ابني / ابنتي مستقبلا إلى صف أعلى.

I also understand that to meet the requirements of the Ministry of Education successfully in the British curriculum it will need to be completed until year 13.

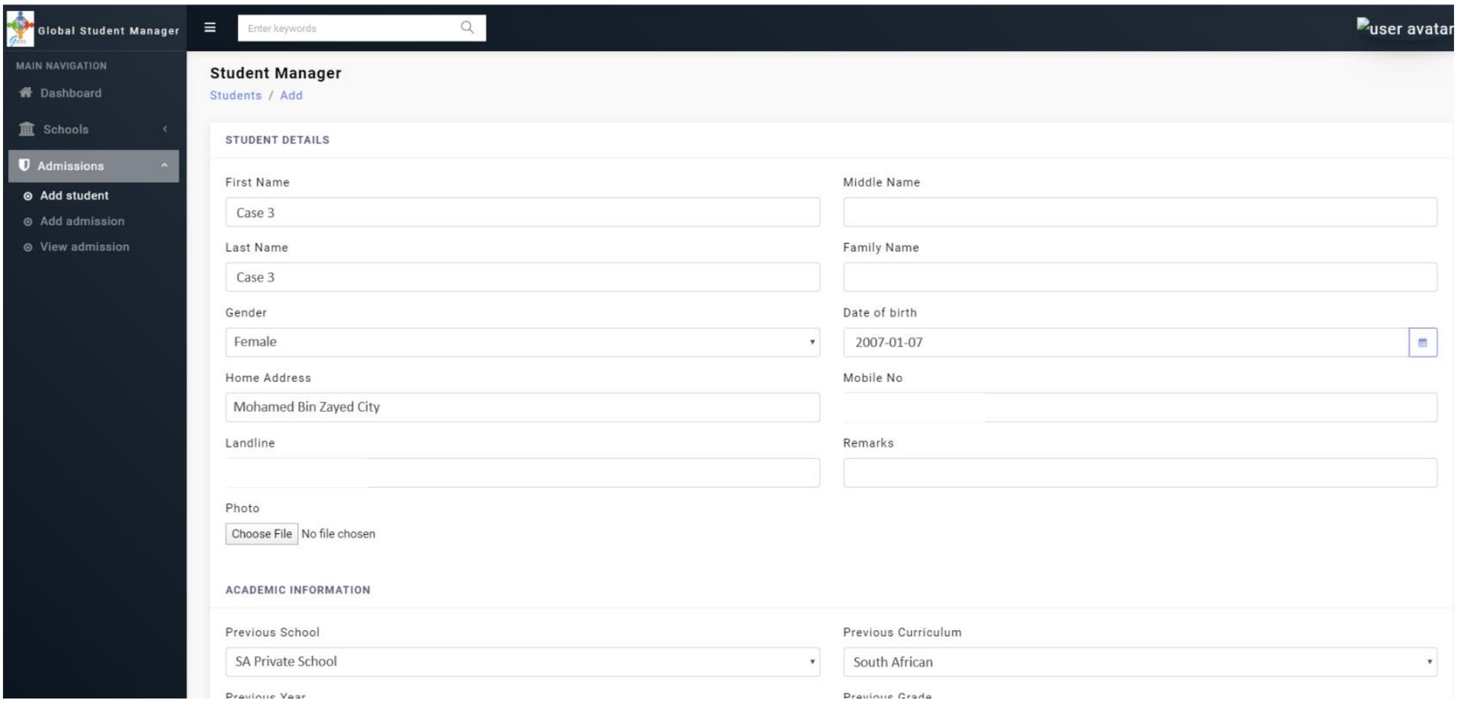
أتفهم أنه لتحقيق متطلبات معادلة وزارة التربية والتعليم بنجاح في المنهاج البريطاني فإنه سيحتاج إلى إكمال حتى السنة 13.

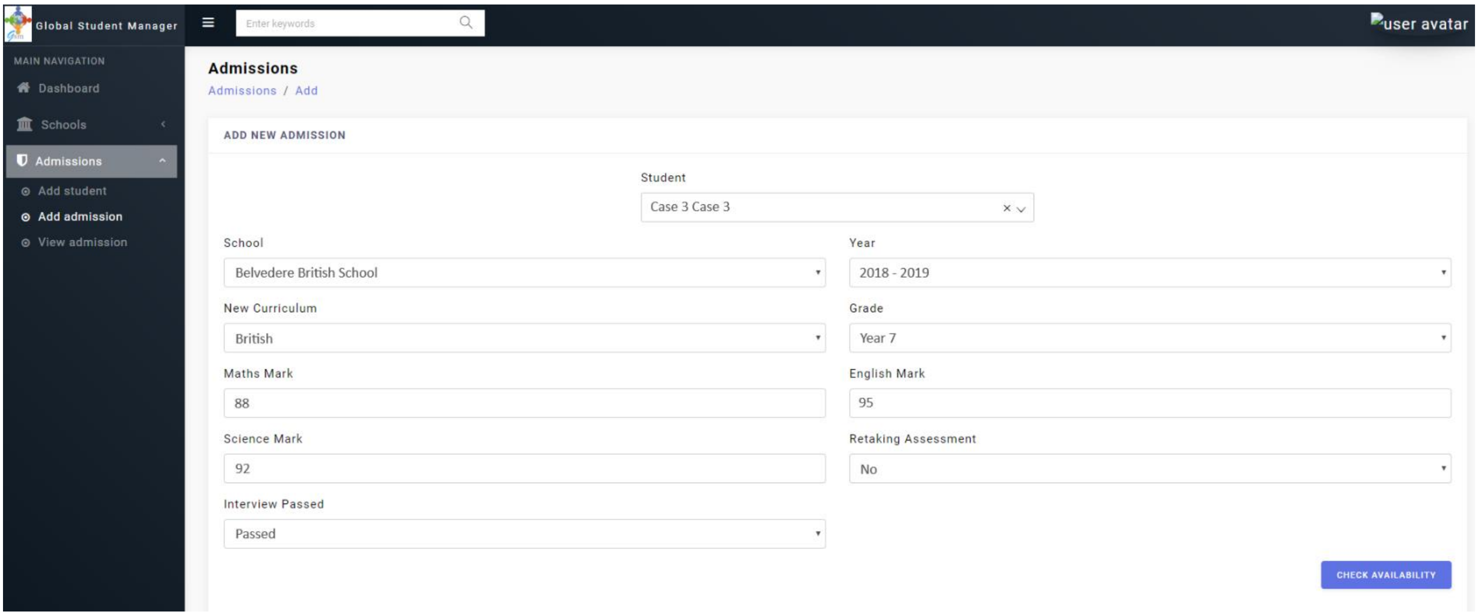
Parent Name:
Signature:
Mobile Number:
Date:

اسم ولي الأمر:
التوقيع:
رقم الموبايل:
التاريخ:

School Principal Decision

Student ID: 53567





BELVEDERE
British School

Parent Undertaking

تعهد ولي الأمر

I hereby confirm that I am in full agreement with School Name...Belvedere....British School Which follows the .British curriculum year system to repeat year...6.. for my child for the academic year 2018/2019

أتعهد بموجب أدناه أنني أوافق كليا مع
مدرسة.. التي تطبق المنهاج
.... نظام ...... بإعادة السنة .... لابني/ابنتي للعام
الدراسي 2019/2018

Student Name: .
Date of birth: ...
eSIS Student ID.........

إسم الطالب:.......
تاريخ الميلاد: ....
رقم الطالب على أيسس: ......

I understand that the normal grade progression would place my child in Year ...6...for the academic year 2018/2019 , however I am requesting that my child repeat Year

أتفهم أن التسلسل الدراسي الصحيح سيضع طفلي في السنة.......... للعام الدراسي 2019/2018 ومع ذلك أطلب أن يعيد أبني السنة .... نظرا للأسباب أدناه:

-1
-2

....6..... Because of the below Reason: -3

1- We emigrated from South africa IN February
2- and they finished the school year IN December
3-

I confirm that I will not be in a position to approach the Department of Education and Knowledge or the school to promote my child to a higher grade/year in the future.

كما أتعهد أنني لن أطالب دائرة التعليم والمعرفة أو المدرسة بترفيع ابني / ابنتي مستقبلا إلى صف أعلى.

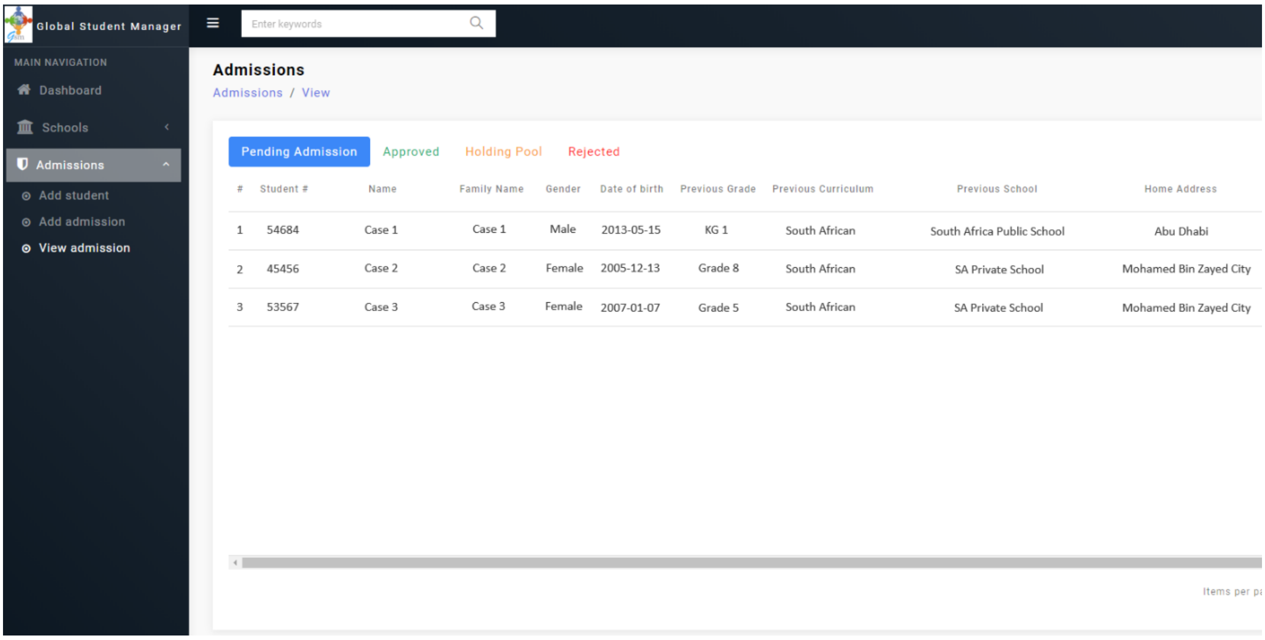
I also understand that to meet the requirements of the Ministry of Education successfully in the British curriculum it will need to be completed until year 13.

أتفهم أنه لتحقيق متطلبات معادلة وزارة التربية والتعليم بنجاح في المنهاج البريطاني فإنه سيحتاج إلى إكمال حتى السنة 13.

Parent Name:
Signature:
Mobile Number:
Date:

اسم ولي الأمر:
التوقيع:
رقم الموبايل:
التاريخ:

School Principal Decision


Global Student Manager
Enter keywords
MAIN NAVIGATION
Dashboard
Schools
Admissions
Add student
Add admission
View admission
Admissions
Admissions / View
Pending Admission
Approved
Holding Pool
Rejected
Student # Name Family Name Gender Date of birth Previous Grade Previous Curriculum Previous School Home Address
1 54684 Case 1 Case 1 Male 2013-05-15 KG 1 South African South Africa Public School Abu Dhabi
2 45456 Case 2 Case 2 Female 2005-12-13 Grade 8 South African SA Private School Mohamed Bin Zayed City
3 53567 Case 3 Case 3 Female 2007-01-07 Grade 5 South African SA Private School Mohamed Bin Zayed City
Items per pa