



**Early detection of academic weakness
in students using artificial intelligence tools**

Hissa Naif Al-Otaibi

252001788

Master's student in Artificial Intelligence

MIDOCEAN UNIVERSITY

2025

Chapter 1 :Introduction

1.1 Problem Definition

The Kingdom of Saudi Arabia is keen on education and its importance, and seeks to develop the educational process and provide a suitable environment for students. Students are considered an important focus in the educational process and must be supported, and any problems that hinder their learning must be resolved. One of the most prominent problems is poor or delayed academic achievement. Some researchers have defined academic achievement as

“all performance by students in various school subjects that can be measured by test scores, teacher assessments, or both.” [Hadi Shaalan Rabie and Ismail Muhammad al-Ghoul, *The Educational Counselor and His Active Role in Solving Student Problems*, Jordan, Dar Al-Thaqafa, 2006]

Academic weakness is a decline in a student's academic performance and can be caused by various social or psychological factors. It is necessary to address and highlight this problem, identify it early, and treat it so that the student can catch up with the rest of the students and prevent them from falling behind.

According to a study by Casillas, A., Robbins, S. B., Allen, J., Kuo, Y. L., Hanson, M. A., & Schmeiser, C. (2012).

which studied the school environment and included 4,660 students from 24 schools, the results indicate that the strongest indicator is the prediction of previous grades, projects, and activities, in addition to psychological and social factors. From these results, we emphasize the importance of early and effective assessment based on the two previous points in order to identify risks and intervene to improve academic performance.

A study by Laetitia Cassells (2012) indicated that students at risk of academic failure benefited more from the early detection system and additional lessons than students who were not subject to the system, achieving cumulative grades that were 8% higher.

1.2 Thesis Objectives

- To review research studies on the causes of student academic weakness, methods of treatment, and ways to enrich high achievers
- To collect and analyze student grade data from the first period of the semester
- To create a tool to help teachers identify students with low grades early on and help them pass the course
- To provide teachers with remedial and enrichment plans

1.3 Thesis Contribution

This work focuses on middle and high school students to help reduce end-of-year academic failure and help teachers identify students at risk of failing a course, develop remedial plans for them, identify high-achieving students, and develop enrichment plans for them, thereby enhancing the quality of the educational process. This is done by creating a tool designed with machine learning models and using real student data to provide interpretable results and take into account future expansion.

1.4 Thesis Organization

Chapter 1 : Introduction and Overview

The first chapter will provide an overview of the study topic, review the problem and objectives, explain how the study contributed to the field of education, and describe the organization of the thesis. reviews traditional methods of machine learning and advanced methods of deep learning. It also explains hybrid models that combine traditional methods with deep learning and reviews quantitative results, research gaps, and opportunities.

Chapter 2 : Background and Related Work

This chapter will provide a comprehensive background on previous studies in the field and identify current trends and their potential for future development.

Chapter 3 : Proposed Approach

Chapter4:Experimental Resultsand Discussion

Chapter 5: Conclusion and Future Work

1.5 Overview

1.5.1 Introduction

Artificial intelligence models play an important role in education and improve and develop the educational process. Artificial intelligence models are divided into several sections, which we will now review.

1.5.2 Classical Machine Learning Methods

I have read studies that use the method and show that the method works well for predictions. However the method shows problems when the method has to handle a lot of data. The problem has caused modern deep learning models to appear. The modern deep learning models help schools and big institutions. Examples of models include:

- **K Nearest Neighbour (KNN)** Classifies data based on the most common points among its K nearest neighbors and is used to classify images, aggregate data, and recognize patterns.
- **The Random Forest** builds a set of decision trees. The Random Forest aggregates the decision trees to produce the classification. The Random Forest uses the classification to classify data that has many variables.
- When I work with **Decision Tree** I watch Decision Tree divide the data into branches based on the feature values. Decision Tree is used in the decision making, the risk assessment and the classification tasks. Decision Tree keeps splitting the data until the leaves are all the same meaning all the cases in a leaf belong, to the same class.
- I have found that **XGBoost** builds the set of decision trees. Each tree in XGBoost is better than the tree and makes fewer errors. I have seen that XGBoost uses regression to improve performance. I use XGBoost for classification, regression and ranking. I have observed that XGBoost has accuracy and efficiency, in educational competitions. XGBoost runs fast. I have used XGBoost on complex data and XGBoost works well.

In general, deep learning models have a high capacity to improve prediction accuracy and analyze complex patterns in educational data, making them effective for supporting decision-making and

developing intelligent educational systems, especially in the context of predicting academic performance and early detection of at-risk students.

.

Of these methods, it is clear that we can use:

- Decision Tree
- KNN
- RF
- XGBoost

to predict student grades early on and identify weaknesses. Data will be collected from several schools in Saudi Arabia over a period of two years, and noise, loss, and unbalanced distribution will be processed to ensure the effectiveness of the model and the accuracy and comprehensiveness of the predictions to help teachers identify students who are prone to failure and intervene to address their weaknesses. Then comes the role of measuring the impact on students following the intervention after the prediction and determining the extent to which the study's objective has been achieved

1.5.3 Use of Real-world Datasets

Collect student grade data from the first and second semesters of the previous two years, along with data from the first semester of this year, including student information (midterm exams, participation grades, assignments, and projects) from several schools, in addition to a database from Kaggle and Zenodo. It is necessary to address the difference in grade distribution between databases and analyze grades to identify student patterns and understand the relationship between them. Noise, missing data, and unbalanced distribution will be addressed to ensure the effectiveness of the model and the accuracy and comprehensiveness of predictions to help teachers identify students who are at risk of failure and intervene to address their weaknesses. Next comes the role of measuring the impact on students after the intervention following the prediction and determining the extent to which the study's objective has been achieved.

1.5.4 Quantitative Results Summary

References	Model-Used)	Data Set	Results
Khan, I., Ahmed, M., & Ali, S. (2021).	• k-Nearest Neighbors (K-NN) • Multilayer Perceptron (ANN) • Naïve Bayes • Decision Tree (RepTree)	collected from the University of Amman over three semesters, covering 151 cases and ten predictive features of student performance	MCC (Matthews Correlatio Coefficient) (K-NN)=0.517 (ANN)=0.585 Naïve Bayes=0.599 Decision Tree=0.637
Deo, R. C., Yaseen, Z. M., Al-Ansari, N., Nguyen-Huy, T., Langlands, T. A. M., & Galligan, L. (2020)	• ELM model • Random Forest (RF) • Volterra model	Six years of student performance data from the University of Southern Queensland, Australia Included several in-person and distance learning courses	prediction error ELM model :0,74% Random Forest (RF) :3.12% Volterra model :1.06%
Zeineddine, H., Braendle, U., & Farah, A. (2021)	• SVM • K-NN • Random Forest • Gradient Boosting	The study used university student data comprising 461 cases for a specific period between 2016 and 2021	Relative Prediction Error • SVM: 3.5% • K-NN:8.0% • Random Forest:6.0% • Gradient Boosting:4.5%
Nimy, E., Mosia, M., & Chibaya, C. (2023)	• Logistic Regression • Support Vector Machine (SVM) • Decision Tree • Random Forest • Naive Bayes	A university in South Africa used student academic data such as tests and assignments, as well as their interaction on the platform, such as the amount of time they spent on the platform, and demographic data.	• Logistic Regression:11.6% • (SVM):10.5% • Decision Tree:12.7% • Random Forest:12% • Naive Bayes :7.19%
Lau, E. T., Sun, L., & Yang, Q. (2019)	• ANN neural networks	The study included 1,000 cases from a Chinese university and three university programs over four years and used 11 variables	ANN neural networks: 84.8
Mondal, A. & Mukherjee, J. (2018)	• Recurrent Neural Network (RNN) • Artificial Neural Network (ANN) • Deep Neural Network (DNN)	Data from Kaggle was used	prediction error • (RNN):85% • (ANN):82% (DNN):83%
Embarak, O. H., & Hawarna, S. (2023).	• Artificial Neural Networks (ANN) • Random Forest • XGBoost algorithm • Support Vector Machine (SVM)	The study collected data from 450 university students from different universities. The data included academic, behavioral, and psychological data	Accuracy • ANN:82.22% • Random Forest:84.44% • XGBoost algorithm:87.78% • (SVM):85%
Melo, E. C., & Souza, F. S. H. (2023).	• Semi-Supervised Learning • Logistic Regression • K-Nearest Neighbors (KNN) • Decision Tree • XGBoost	From an educational institution in Brazil for one year, including • 5,915 cases of dropouts • 2,096 cases of graduates	Accuracy • Semi-Supervised Learning:90% • Logistic Regression:85.5% • (KNN):87.68% • Decision Tree:88.05% • XGBoost 92.05%

1.5 .5 Research Gaps and Opportunities

Connections:

- The studies relied on real student data.
- The studies were conducted within a specific scope within a single educational institution.
- The researchers used more than one algorithm in the studies and selected the best one.
- The models showed remarkable accuracy in the results with a negligible error rate.

Research gap:

- The need to expand the scope of the study so as not to affect the generalizability of the results.
- Adding the largest possible segment of students to contribute to increasing the efficiency and accuracy of the model.
- Adding multiple features, different educational materials, and varying educational levels.
- Need to focus on understanding the model's recommendations to help teachers make decisions
- Need to study the practical application of prediction to guide real educational interventions, determine the effectiveness of the model, and measure its impact on students

Chapter 2: Background and Related Work

Research has focused on using artificial intelligence and machine learning algorithms to predict student outcomes in educational settings. The goal is to improve understanding of student academic performance, develop systems for early detection of students at risk of academic failure or dropout, and implement appropriate preventive measures. Among the most prominent algorithms used in this field are artificial neural networks (ANNs), semi-supervised learning, random forests, and probabilistic learning. Research and studies on this topic have been diverse.

A group of these studies focused on predicting student performance and early detection of at-risk students using traditional artificial intelligence techniques. Studies such as **Khan et al. (2021)** collected data from the University of Amman over three semesters, covering 151 cases and ten predictive characteristics of student performance. The study identified the most important factors affecting performance as GPA, exam scores, and attendance, and determined that students with a GPA of 2.75 or higher showed better results and performance than those who obtained averages below 2.75. The study used several models, the most accurate and best in performance being the decision tree. There was also a study by **Lau et al. (2019)** that included 1,000 cases from a Chinese university and three university programs over four years and used 11 variables. The results improved over the years, and the study used the ANN model, which gave results with an accuracy of 84.8%. We also have a study by **Sukhbaatar et al. (2019)**. Data was collected from Kumamoto University as part of a compulsory course for second-year students. The number of students reached 1,110 over six semesters from 2012 to 2017. A multi-layer neural network was used, and the model identified 25% of students at risk of failure after the first short test and after

the second test. The midterm test results improved, and 65% of students at risk of failure were identified due to increased data and, consequently, increased model accuracy.

This group focused on students' previous grades, attendance, and tests. One of the most notable findings of this research is that neural networks and decision trees outperformed other models, with prediction accuracy ranging from 86% to 91%, enabling educational institutions to provide remedial plans for struggling students, such as extra lessons and early guidance.

In contrast, another group of studies used probabilistic and advanced models to achieve more accurate early predictions, such as the studies by **Nimy et al. (2023)** used academic data from students at a university in South Africa, such as tests and assignments, in addition to their interaction on the platform, such as the amount of time they spent on the platform and demographic data. The researchers used probabilistic machine learning and AutoML techniques to measure the probability of academic failure at each stage of the semester, rather than just giving a final result. The results showed that these models supported the reliability of early prediction, with an accuracy of 92%.

A bunch of studies also used advanced techniques like recurrent neural networks (RNN) and semi-supervised learning, like in the studies by **Mondal and Mukherjee (2018)** and **Melo and Souza (2023)**, which used data from an educational institution in Brazil for one year, including 5,915 dropouts and 2,096 graduates. Researchers **Mondal, A. & Mukherjee** data from Kaggle, where RNN was successful in extracting temporal patterns, and these models were used to analyze educational data. The models in this group showed great effectiveness in tracking changes in student performance over time and early prediction of academic failure, proving their usefulness in long-term early intervention programs.

The latter group, such as the studies **by Deo et al. (2020)**, which relied on six years of student performance data from the University of Southern Queensland, Australia, included several in-person and distance learning courses, and the ELM model outperformed the other models in predicting student grades and identifying those at risk of failure with a very low error rate. It was faster in terms of computational efficiency and pattern capture in continuous assessment data and helps the model improve the quality of teaching and early support. For the study **by Zeineddine et al. (2021)** used data from 461 university students for a specific period between 2016 and 2021. It focused on developing high-performance models using AutoML to predict student grades in specific courses and used student data available before the start of the academic program. It showed high prediction accuracy with low errors, outperforming traditional models such as random forests or traditional neural networks, with an accuracy of 83%.

Conclusion

The literature supports the use of artificial intelligence and shows its success in identifying students at risk of failing with high accuracy using different models. Based on these results, we find that the results of the models differ between studies, but the most accurate was XGBoost with 92%, while the most commonly used models are recurrent neural networks (RNN) and Random Forest. Therefore, the appropriate models for the study will be selected and trained on real data after processing and improving feature selection to give higher quality and credibility results, thus contributing academically and scientifically in this field.

Chapter 3 : Proposed Approach

The curriculum is based on rigorous data analysis and the optimization of machine learning models to obtain realistic and accurate results, which are used to identify at an early stage students at risk of academic failure.

3.1 Data Collection and Description

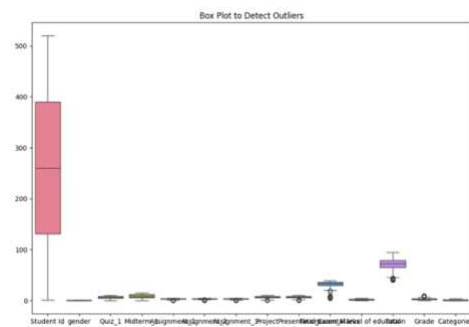
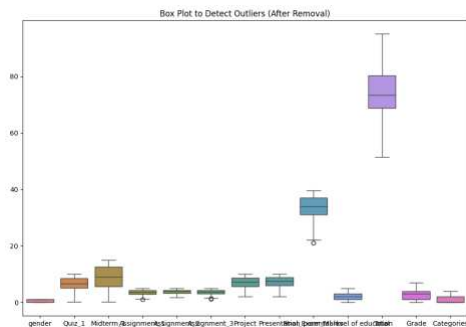
A dataset from Kaggle related to student academic performance was used. It contains students' grades on tests, assignments, projects, and presentations, as well as other data such as gender, parents' education level, and final exam scores. The data was categorized to analyze student performance and classify them into groups indicating each student's level and predicting categories of academic weakness

3.2 Data Preprocessing

After collecting the data, I applied several data processing steps, such as

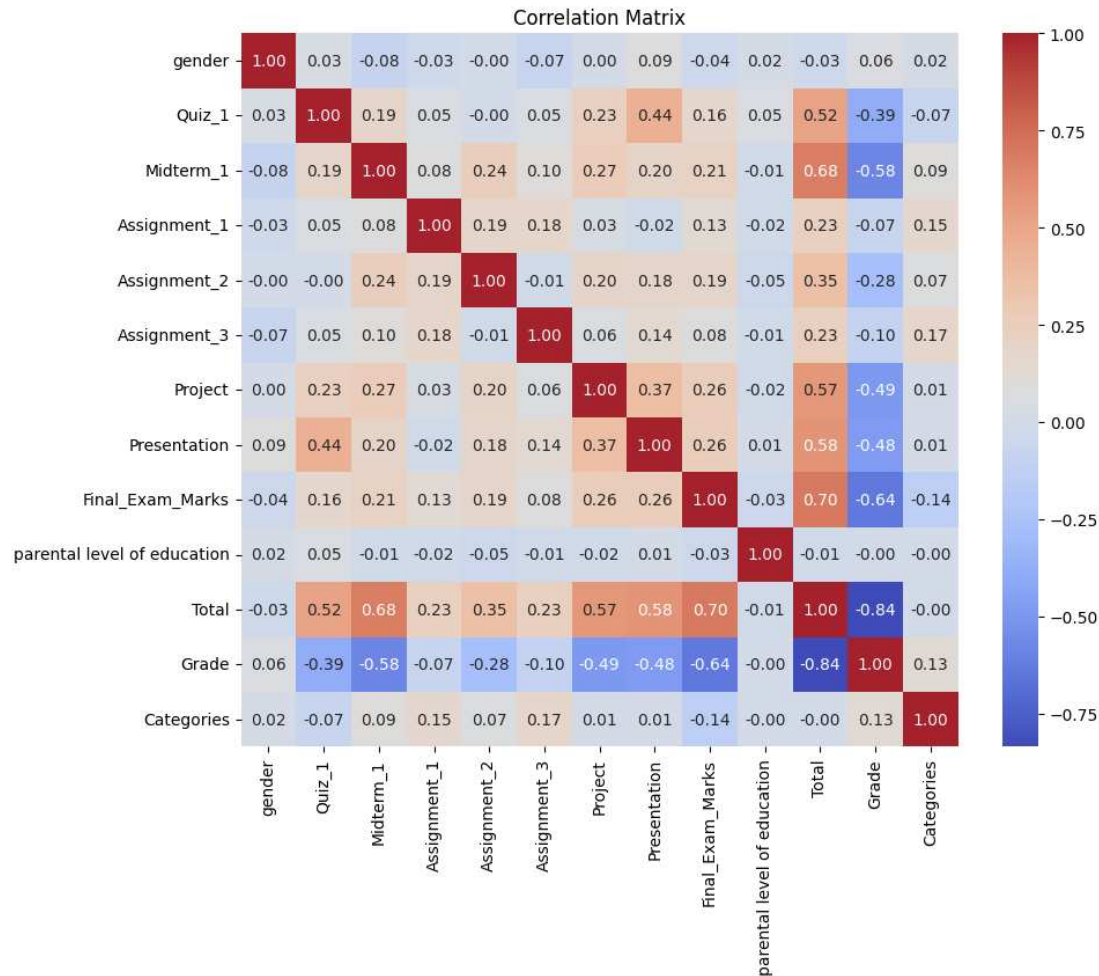
- **replacing commas with periods:** The dataset contains decimal commas that can only be processed after converting them to periods
- **Converting text columns to numbers:** I found some columns such as gender, parental level of education, Grade, and Categories, all of which contained text data that was converted to numerical data appropriate for their type
- **Handling missing data**
- **Missing values were replaced with the mean values**

- Removing some columns:** To achieve accurate results, we had to remove certain columns such as Student Id because it provides no benefit to the predictions. The Grade and Total columns were also removed because they resulted in extremely high prediction rates, leading to unrealistic results; the model interpreted them as a reflection of the student's level, with accuracy rates reaching 100% for some models. This adjustment aims to prevent overfitting and ensure the data is balanced and well-suited for model training
- Detecting outliers:** Outliers were detected and visualized in the dataset using the interquartile range and box plots. These values were handled using the Interquartile Range (IQR) to identify values outside acceptable limits, which were then removed to ensure the accuracy of the data used.



2.3 Feature Engineering

The features available for training were optimized based on the original data, such as extracting interactive features between different test scores.



3.4 Feature Selection

Features with a strong influence on the prediction of academic underachievement were selected by examining the previous correlation matrix, analyzing the relationships between features, choosing those most closely associated with academic underachievement, and removing features that cause overfitting—such as Grade and Total—to improve class distribution.

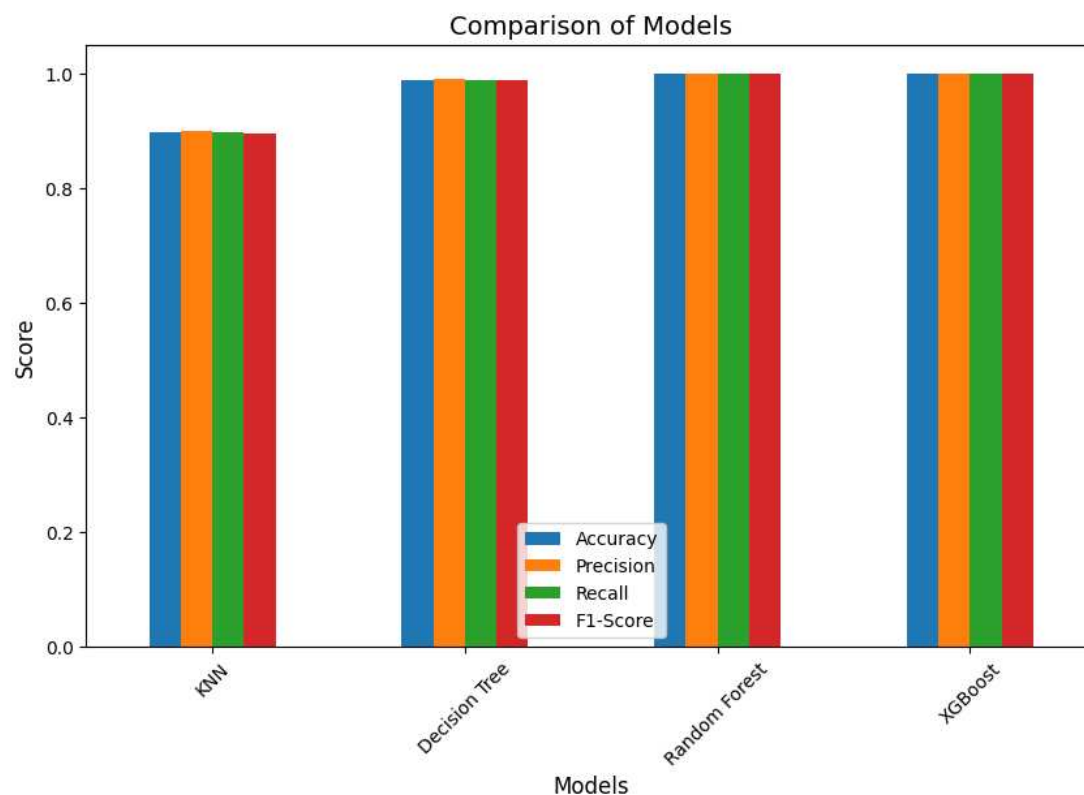
3.5 Model Selection and Optimization

The machine learning models used are suitable for the task of predicting academic underachievement, including KNN, Decision Tree, Random Forest, and XGBoost

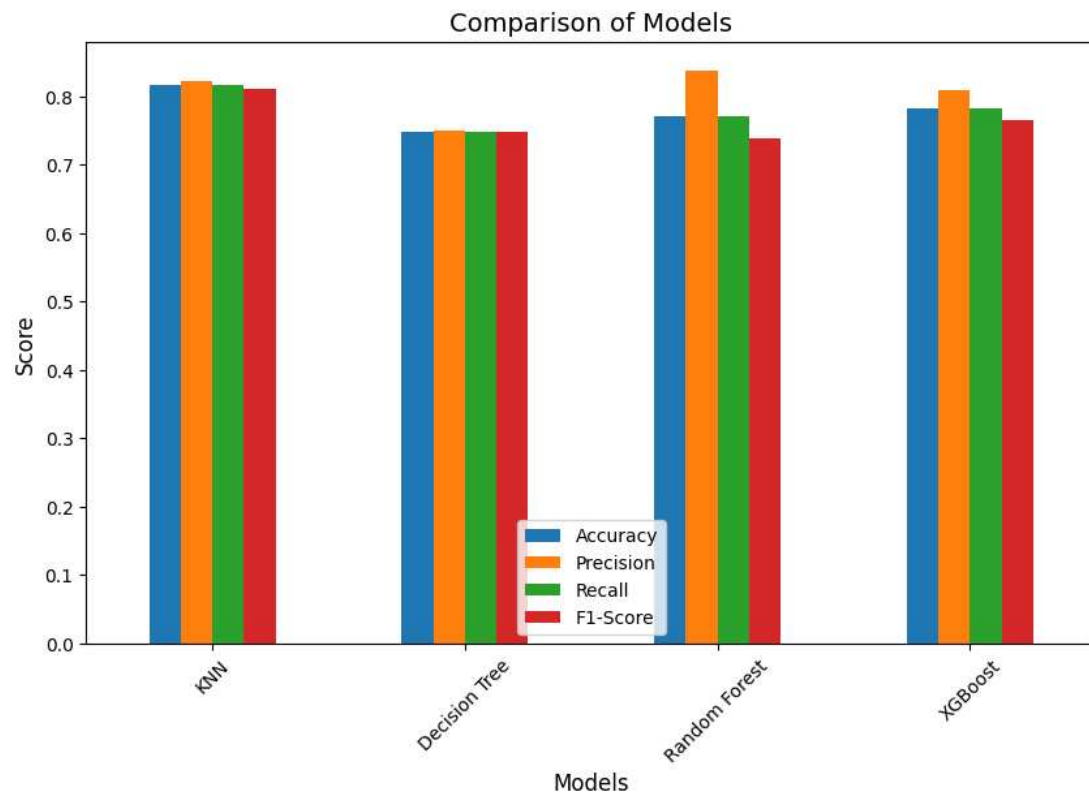
They were optimized using GridSearchCV to obtain the best parameters, such as the number of trees in Random Forest and XGBoost and the tree depth in Decision Tree. We also used optimization techniques such as cross-validation to reduce overfitting and ensure that the models are capable of good generalization

3.6 Model Evaluation

The results were evaluated using various evaluation metrics such as Accuracy, Precision, Recall, and F1 score, using two methods. The first method included all features present in the dataset, and the results were unrealistic and very high due to the presence of Grade and Total, with accuracy reaching 100%, indicating overfitting

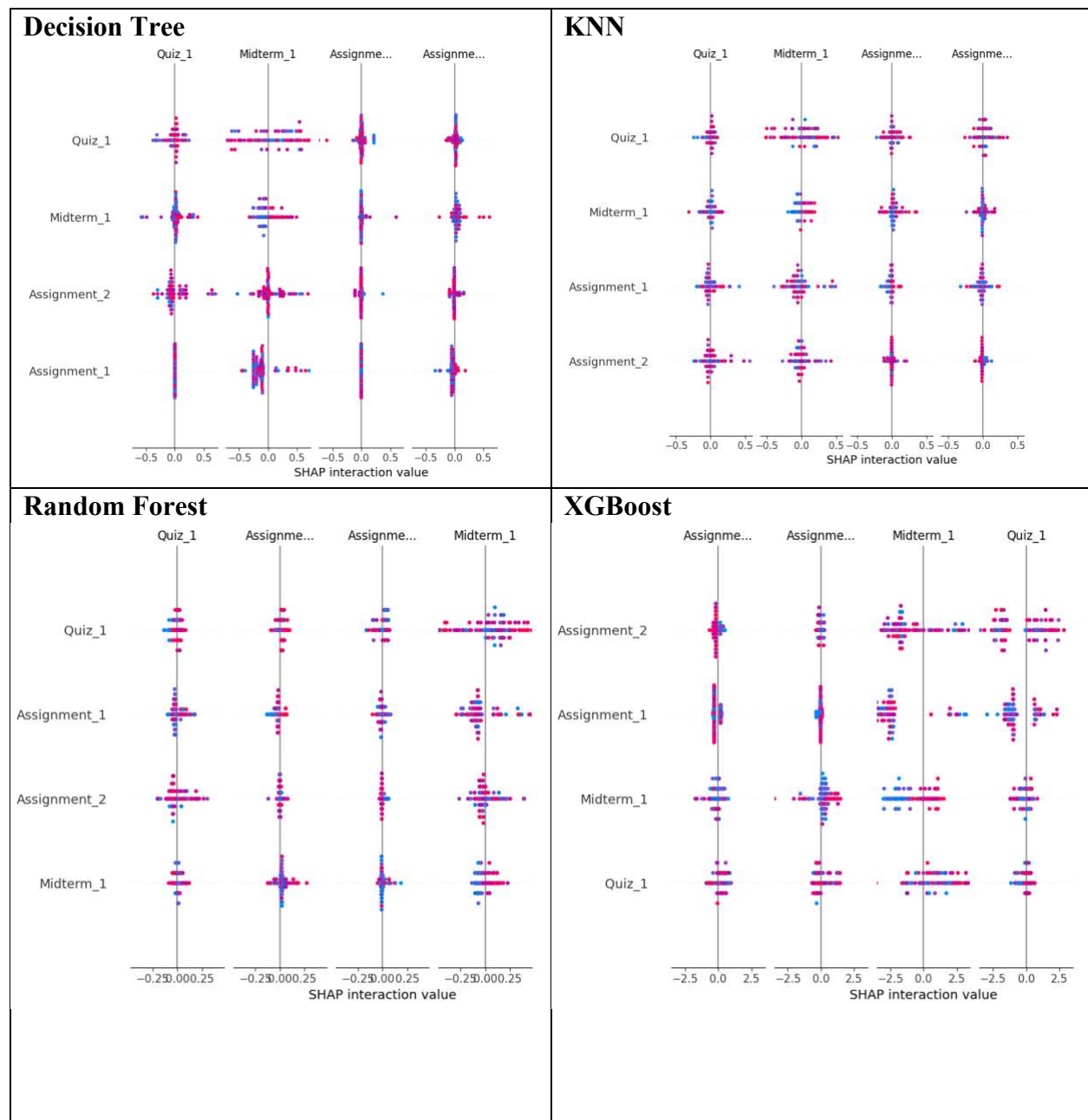


Therefore, the features causing high accuracy were removed because the models treated them as redundant, and the results became more realistic



3.7 Visualization and Insights

After using SHAP diagrams, it becomes clear that certain features have a significant impact on the final decisions regarding models, such as assignments and projects



Chapter 4

Experimental Results and Discussion

4.1 Experimental Setup

All experiments were implemented in Python using several libraries:

- pandas
- numpy
- matplotlib.pyplot
- seaborn

The baseline dataset consists of 521 records. Outliers were handled, and text data was converted to numerical data. The dataset contained 14 columns; columns that caused overfitting issues were removed, and the data was split into 80% for training and 20% for testing

5.2 Baseline Model Performance

All features were used to train the models initially, and the results showed very high accuracy, indicating overfitting. The performance results were as follows:

(Model)	(Accuracy)	(Precision)	(Recall)	F1 (F1-Score)
KNN	0.90	0.90	0.90	0.89

Decision Tree	0.99	0.99	0.99	0.99
Random Forest	1.00	1.00	1.00	1.00
XGBoost	1.00	1.00	1.00	1.00

4.3 Feature Selection and Model Refinement

We then removed columns that were irrelevant to the prediction process, such as

student ID

, and removed features that caused overfitting, such as

Total: which provides the total score in numerical form, reflecting the student's level or classification in Categories

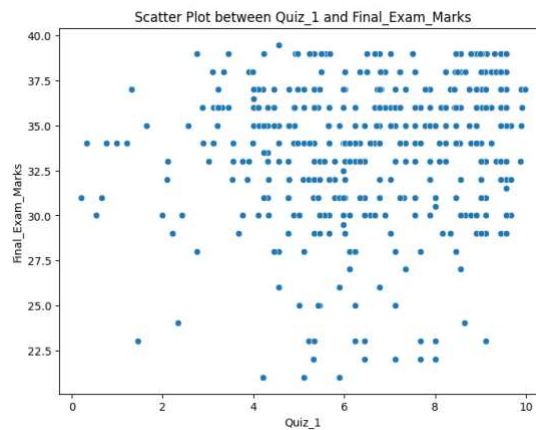
Grade: converts the previous total into letter grades, which also indicate the student's level

Therefore, it was necessary to remove these features, and the results became more realistic. After adjusting them with GridSearchCV, the performance metrics appeared as follows

(Model)	(Accuracy)	(Precision)	(Recall)	F1 (F1-Score)
KNN	0.82	0.83	0.82	0.81
Decision Tree	0.75	0.75	0.75	0.75
Random Forest	0.77	0.84	0.77	0.74
XGBoost	0.78	0.81	0.78	0.76

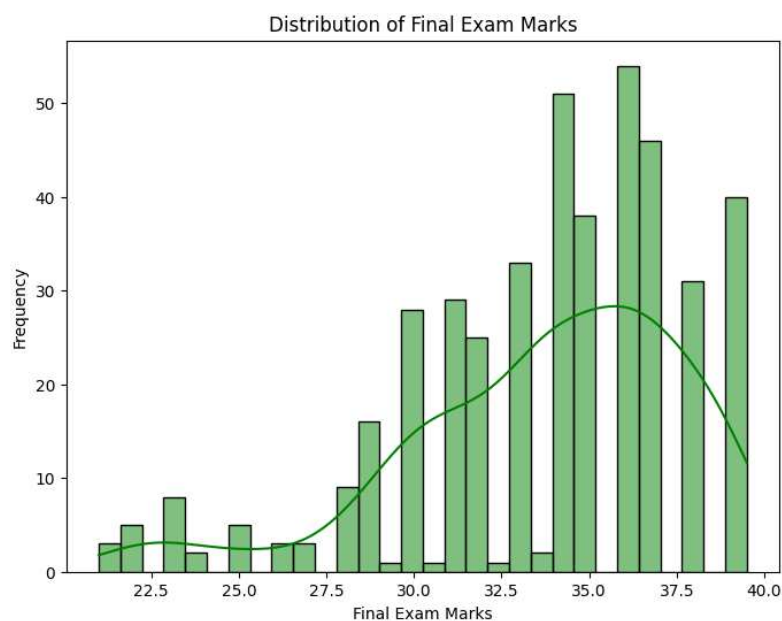
5.4 The Relationship Between Quiz_1 and Final Exams

The scores indicate that the correlation between them is very weak; it is not necessary for a student to receive a high score on the midterm exam in order to receive a high score on the final exam.



4.5 Analysis of Final Exam Scores

We note that most students achieved good results and demonstrated a high level of performance, while a small number received low scores; most scores fell between 35 and 37.5



4.6 Discussion

The use of artificial intelligence in predicting student performance is important for improving their final results and identifying underperforming students early on so that teachers can intervene and address their weaknesses. The models yielded good results after a careful selection of appropriate features. The most significant findings were as follows:

- A student's performance on projects and presentations has a strong correlation with test scores
- Students who performed well on the midterm exam achieved high scores on the final exam
- Simplifying the model improved results by removing several features and focusing on those most influential on student performance
- The KNN model showed more balanced and stable results and achieved the highest accuracy value, reaching 0.82
- The RF model achieved the highest Precision value, meaning it is best at reducing false positives
- The model with the lowest performance is DT compared to the other models

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Early prediction of student performance is one of the most important factors in addressing academic underachievement, and artificial intelligence has proven its effectiveness and accuracy in identifying this group, while the models demonstrated varying degrees of accuracy, some of which were quite good. It became clear to us from the data set that students' performance on assignments and projects plays a significant role in improving their grades, as students who are engaged in their work perform well on exams, unlike those who are not. Furthermore, data analysis revealed that students who scored high on the midterm exam were also able to perform well on the final exam, which highlights the importance of closely monitoring students before final exams and ensuring they complete all assignments correctly. Therefore, it is important to incorporate artificial intelligence into education and use it as a tool to assist both teachers and students in achieving good results, reducing academic underperformance among students, and reducing the teacher's workload.

5.2 Future Work

- Use of deep learning techniques: We can expand the scope of the curriculum to incorporate deep learning techniques, such as neural networks, to improve predictions

- Adding new features: We can introduce a set of new features to improve predictions, such as student online behavior, previous years' grades, or interaction with educational platforms
- Model optimization: We can optimize the models using other techniques such as Gradient Boosting and LightGBM to achieve better performance.

References

1. A. Casillas, S. B. Robbins, J. Allen, Y.-L. Kuo, M. A. Hanson, and C. Schmeiser, "Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior," *Journal of Educational Psychology*, vol. 104, no. 2, pp. 407-420, 2012, doi: 10.1037/a0027180.
2. L. Cassells, "The effectiveness of early identification of 'at risk' students in higher education institutions," University of Pretoria, 2012
- 3 - [Hadi Shaalan Rabie and Ismail Muhammad al-Ghoul, *The Educational Counselor and His Active Role in Solving Student Problems*, Jordan, Dar Al-Thaqafa, 2006]
- 4 - S. M. Lange and B. Thompson, "Early identification and interventions for children at risk for academic failure," *International Journal of Special Education*, vol. 21, no. 3, pp. 108–117, 2006.
- 5 - A. Rashid, R. Yasmeen, R. Ahmed, and K. Jawed, "Factors leading to the academic failure of undergraduate medical students
- 6- Khan, I., et al., An artificial intelligence approach to monitor student performance and devise preventive measures. *Smart Learning Environments*, 2021. 8(1): p. 1-18.
- 7- Deo, R.C., et al., Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. *IEEE Access*, 2020.
- 8- Zeineddine, H., U. Braendle, and A. Farah, Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, 2021. 89: p. 106903
- 9- Hashim, A.S., W.A. Awadh, and A.K. Hamoud. Student performance prediction model based on supervised machine learning algorithms. in *IOP conference series: materials science and engineering*. 2020. IOP Publishing.
- 10- Lau, E.T., L. Sun, and Q. Yang, Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 2019.
- 11- Pallathadka, H., et al., Classification and prediction of student performance data using various
- 12-machine learning algorithms. *Materials today: proceedings*, 2023.
- Sukhbaatar, O., T. Usagawa, and L. Choimaa, An artificial neural network based early prediction of failure-prone students in blended learning course. *International Journal of Emerging Technologies in Learning (iJET)*, 2019. 14(19)
- 13- Burkov, Andriy. *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019. ISBN: 978-1999579500.

14- Melo, E. C., & Souza, F. S. H. (2023). *Improving the prediction of school dropout with the support of the semi-supervised learning approach*. *iSys: Brazilian Journal of Information Systems*, 16(1). <https://doi.org/10.5753/isys.2023.2852>

15- Mondal, A. and J. Mukherjee, An Approach to predict a student's academic performance using Recurrent Neural Network (RNN). *Int. J. Comput. Appl*, 2018

16- **Nimy, E., Mosia, M., & Chibaya, C. (2023).** Identifying At-Risk Students for Early Intervention — A Probabilistic Machine Learning Approach. *Applied Sciences*, 13(6), 3869. <https://doi.org/10.3390/app13063869>