



Predictive Modeling for Identifying Early Warning Signs of Underperformance in Vocational Education

D. S. S Hettiarachchi
(Reg. No.: MS23004368)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Mr. S.M.B. Harshanath

Approved for MSc. Research Project:

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.



Sign:

D. S. S. Hettiarachchi

Date:2025/01/16.....

ABSTRACT

Predictive Modeling for Identifying Early Warning Signs of Underperformance in Vocational Education

D.S.S. Hettiarachchi

MSc. in Information Technology

Supervisor: Mr. S.M.B. Harshanath

December 2024

Vocational education is vital for developing a skilled workforce that meets national economic and social demands. However, high dropout rates and underperformance challenge this sector, often due to limited monitoring tools that capture early warning signs in practical, skill-based environments. This study addresses these challenges by developing a predictive model to identify at-risk students in vocational education, utilizing several machine learning techniques including Logistic Regression, Decision Trees, Random Forest, and Neural Networks to provide actionable insights for educators and administrators. Among these, Random Forest emerged as the most practical and effective model, especially after applying transfer learning to adapt academic datasets to vocational contexts. The research addresses key questions: identifying early indicators of underperformance, evaluating predictive techniques suited to vocational settings, and establishing model accuracy and reliability for real-world application. By analyzing indicators such as attendance, assessment scores, and student engagement, the model enables early interventions, contributing to improved retention and academic outcomes. Integrated into a Training Management System (TMS), this predictive model demonstrates a practical approach to supporting student success. This research contributes to educational data mining by showcasing how predictive analytics and transfer learning can enhance vocational education, thereby strengthening workforce readiness and employability.

ACKNOWLEDGEMENT

I extend my heartfelt gratitude to the Sri Lanka Institute of Information Technology (SLIIT) for offering this MSc program, which has been an extremely valuable opportunity for me. Each subject's comprehensive curriculum and depth have significantly enriched my knowledge and skills. This program has strengthened my expertise and empowered me to confidently pursue my professional and academic goals. I am deeply appreciative of the dedication of the faculty and the high standards set by the Institute, making this course a truly rewarding experience for all of us.

During my time at SLIIT, I have had the privilege of working with remarkable advisors, each contributing unique perspectives and insights. I am especially grateful to my supervisor, Mr. H.M.B. Harshanath, for his invaluable mentorship throughout my research journey. His thoughtful guidance, constant constructive feedback on my ideas and writing, access to his network of contacts, and the freedom to pursue my projects have profoundly shaped my work.

I am also grateful to my lecturers at SLIIT, who have guided us with a balanced approach—challenging us to excel under time constraints while also knowing when to ease the pressure. Their thoughtful guidance and engaging teaching methods have greatly enriched both my academic and personal growth.

Furthermore, I extend my appreciation to Dr. Shanthikumar Hettiarachchi, whose mentorship has been a beacon of guidance throughout this endeavour. I would also like to express my deepest gratitude to my parents and family, who have been my steadfast supporters through both good times and challenging moments. Their encouragement and belief in me have made this accomplishment possible.

Finally, I am truly thankful for the support of my friends and colleagues, near and far. Their camaraderie and kindness have been a source of strength, sharing in both the joys and challenges of this journey.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
List of Figures	ix
List of Tables.....	x
Chapter 1 Introduction.....	1
1.1 Background.....	1
1.1.1 History and Evaluation of Vocational Education	2
1.1.2 Key Concepts	3
1.2 Research Objectives	5
1.3 Research Questions.....	5
1.4 Scope of the Study	6
1.5 Significance of the Study	6
1.6 Thesis Structure.....	7
Chapter 2 Literature Review	8
2.1 Introduction to Vocational Education.....	8
2.1.1 An Introduction to TVEC in Sri Lanka	9
2.2 Factors Influencing Student Performance in Vocational Education	11
2.2.1 Attendance and Engagement Self Efficacy	11
2.2.2 Assessment Scores	11
2.2.3 Socio-Economic Background and Learning Environment	12
2.3 Challenges in Predicting Underperformance in Vocational Education.....	12
2.3.1 Data Availability and Quality.....	12
2.3.2 Differences between Academic and Vocational Datasets.....	12
2.3.3 Existing Early Warning Systems and Predictive Models in Education	13
2.4 Transfer Learning in Predictive Modeling.....	13
2.5 Training Management System	15
2.5.1 Evolution of TMS.....	15
2.5.2 Integrated Systems and Automation in Training Management.....	15
2.5.3 AI-Driven Systems in Training Management.....	16
2.5.4 Practical Training and Competency-Based Learning Systems	16
2.6 Gap Identification.....	16
2.6.1 Gaps in Predictive Modeling for Vocational Education	16
2.6.2 Lack of Domain-Specific Data	16

2.6.3 Limited Application of Transfer Learning.....	17
2.6.4 Underutilization of Self-Supervised and Semi-Supervised Learning	17
2.6.5 Tailoring Predictive Models to Vocational Education.....	17
2.6.6 Integration of AI-Driven Predictive Models with Personalized Learning	18
2.7 Conceptual Framework	18
Chapter 3 Methodology.....	19
3.1 Research Design	19
3.1.1 Research Strategy	19
3.1.2 Data Source Selection.....	20
3.1.3 Data Clearing and Preparation.....	20
3.1.4 Feature Selection and Engineering	20
3.1.5 Modelling Techniques.....	21
3.1.6 Ethical and Data Privacy Considerations.....	21
3.1.7 Limitations of the Design	21
3.2 Data Collection Methods.....	22
3.2.1 Data Sources.....	22
3.2.2 OULAD Dataset	22
3.2.3 Key Attributes of the Dataset	24
3.2.4 Institutional Dataset.....	26
3.3 Data Preprocessing	27
3.3.1 Load and Explore Datasets.....	27
3.3.2 Data Cleaning.....	28
3.3.3 Create Features	29
3.3.4 Prepare and Merge Course Data.....	29
3.3.5 Data Assessment Process.....	29
3.3.6 Prepare and Merge VLE Data	30
3.3.7 Final Dataset Aggregation	30
3.3.8 Validate the Merged Dataset	30
3.3.9 Final Adjustments.....	31
3.3.10 Store or Use the Prepared Dataset	31
3.4 Feature Selection.....	32
3.5 Feature Engineering.....	33
3.6 Student Engagement.....	33
3.6.1 Calculating Frequency-Based Engagement	33
3.6.2 Calculating Timeliness-Based Engagement	35
3.6.3 Calculating Consistency Engagement	36
3.6.4 Calculating Engagement Diversity	37

3.6.5 Aggregating Engagement Model Training	37
3.7 Final Dataset.....	38
3.8 Data Splitting	39
3.9 Model Development.....	39
3.10 Transfer Learning Techniques	42
3.10.1 Evaluate the Pre-trained Model on the Original Dataset	44
3.10.2 Identify Common Features between Datasets.....	44
3.10.3 Adapt Feature Representations for Vocational Context	44
3.10.4 Freeze Layers in the Pre-trained Model	44
3.10.5 Fine-Tune Specific Layers for Vocational Data	45
3.10.6 Use Transfer Learning Techniques to Enhance Adaptation	45
3.10.7 Evaluate the Fine-Tuned Model on Vocational Data.....	45
3.10.8 Iterate and Adjust as Needed	45
3.11 Transfer Learning Process	46
3.12 Model Evaluation	47
3.13 Implementation.....	48
3.13.1 User Management System.....	49
3.13.2 Predicting System (Machine Learning Model).....	49
3.13.3 Data Flow and Prediction Process	50
Chapter 4 Results	53
4.1 Overview of Results.....	53
4.1.1 Model Performance	53
4.2 Comparative Analysis	57
4.2.1 Before Applying the Transfer Learning.....	57
4.2.2 After Applying the Transfer Learning.....	58
Chapter 5 Discussion.....	66
5.1 Interpretation of Results.....	66
5.1.1 Model Performance Comparison Before and After Transfer Learning	66
5.1.2 Effectiveness of Transfer Learning	67
5.1.3 Feature Importance Analysis	67
5.1.4 ROC Curve Analysis.....	68
5.1.5 Best Algorithm for Predicting Outcomes	68
5.2 Implications for Vocational Education.....	69
5.3 Integration with TMS.....	69
5.4 Comparison with Existing Research	70
Chapter 6 Conclusion	72
6.1 Summary of Key Findings.....	72

6.2 Contributions to Knowledge	72
6.3 Recommendations for Future Research	74
6.3.1 Expanding Dataset Diversity and Granularity	74
6.3.2 Integration of External Factors	74
6.3.3 Longitudinal Analysis for Predictive Consistency.....	74
6.3.4 Combining Multiple Influential Factors	75
6.3.5 Developing Real-Time Intervention Models	75
6.3.6 Empirical Validation and Scalability	75
6.3.7 Policy and Practical Application in Vocational Education	75
6.3.8 Integration of AI in Vocational Training.....	76
6.3.9 Addressing Ethical and Bias Concerns in AI Systems	76
6.3.10 Long-Term Impact of AI-Driven Systems in Vocational Education	76
6.3.11 Scalability and Real-World Implementation	76
6.3.12 Ethical Considerations.....	77
6.3.13 Final Thoughts.....	77
References	78

List of Figures

Figure 3.1 ERD – Entity Relationship Diagram of OLUAD Dataset.....	26
Figure 3.2 Structure of the TMS.....	48
Figure 4.1 Feature Importance	54
Figure 4.2 Feature Importance Score (Mutlual Information).....	55
Figure 4.3 Confusion matrix	56
Figure 4.4 Feature Importance for Predicting Pass/Fail in Updated Vocational Dataset.....	57
Figure 4.5 Accuracy Comparison Before and After Transfer Learning	59
Figure 4.6 Precision Comparison Before and After Transfer Learning.....	60
Figure 4.7 F1- Score Comparison Before and After Transfer Learning.....	60
Figure 4.8 Recall Comparison before and after Transfer Learning	61
Figure 4.9 Permutation Feature Importance for Decision Tree	62
Figure 4.10 Permutation Feature Importance for Random Forest	63
Figure 4.11 Permutation Feature Importance for Logistic Regression.....	63
Figure 4.12 Permutation Feature Importance for Neural Network.....	64
Figure 4.13 ROC Curves for Different Models.....	65

List of Tables

Table 3.1 Data Summary Table.....	25
Table 4.1 Eight different models performances	53
Table 4.2 Model Performance before Apply Transfer learning.....	58
Table 4.3 Model Performance before Apply Transfer learning.....	58