

TABLE OF CONTENTS

1	Introduction	1
1.1	Background and Motivation	1
1.2	The MOOC Revolution and India	3
1.3	The Dropout Crisis: Scale and Consequences	5
1.4	Problem Statement	7
1.5	Research Objectives	9
1.6	Scope and Boundaries	10
1.7	Dissertation Organization	12
2	Literature Review	18
2.1	Historical Development of Dropout Research	18
2.2	Early MOOC Studies (2012–2015)	20
2.3	Machine Learning Era (2015–2019)	29
2.4	Deep Learning Approaches (2019–2023)	30
2.5	Theoretical Frameworks in Learning Analytics	30
2.6	Review of Intervention Effectiveness Studies	30
2.7	Gap Analysis and Research Positioning	30
3	Feasibility Study	33
3.1	Technical Feasibility	33
3.2	Operational Feasibility	34
3.3	Economic Feasibility	35
3.4	Schedule Feasibility	35
4	System Design and Architecture	68
4.1	Design Philosophy and Principles	68
4.2	Three-Layer Architecture Overview	71
4.3	Data Layer Design	74
4.4	Processing Layer Design	78
4.5	Presentation Layer Design	83
4.6	Security and Privacy Architecture	87
4.7	Deployment Infrastructure	91
5	Implementation	95
5.1	Technology Stack	95
5.2	Data Collection Module	98
5.3	Data Cleaning and Preprocessing	103

5.4 Feature Engineering Pipeline	107
5.5 Machine Learning Pipeline	116
5.6 REST API Design	126
5.7 Instructor Dashboard Implementation	130
6 Database Design	135
6.1 Schema Design Philosophy	135
6.2 Entity-Relationship Model	138
6.3 Table Definitions	141
6.4 Indexing Strategy	148
6.5 Data Integrity and Backup	151
7 Testing Strategy and Results	154
7.1 Unit Testing	154
7.2 Integration Testing	158
7.3 System and Load Testing	162
7.4 User Acceptance Testing	166
7.5 Security Testing	170
8 Results and Discussion	174
8.1 Dataset Description	174
8.2 Model Performance Comparison	177
8.3 Confusion Matrix Analysis	181
8.4 Temporal Prediction Analysis	184
8.5 Feature Importance Analysis	188
8.6 Dropout Pattern Analysis	192
8.7 Dashboard Usability Evaluation	196
8.8 Comparative Study	200
9 Conclusion and Future Work	204
9.1 Summary of Contributions	204
9.2 Key Findings	207
9.3 Directions for Future Research	211
10 Limitations of the Study	217
— References	224
— Appendices	231

List of Figures

Figure No.	Caption	Page No.
Figure 2.1	Chronological Timeline of Dropout Prediction Research (2012–2024)	28
Figure 2.2	Taxonomy of Machine Learning Approaches Used in Prior MOOC Studies	32
Figure 2.3	Comparison of Algorithm Families: Accuracy vs Interpretability Tradeoff	38
Figure 3.1	Feasibility Assessment Radar Chart	67
Figure 4.1	Three-Layer System Architecture Overview Diagram	72
Figure 4.2	Data Flow Diagram — Complete Prediction Pipeline	76
Figure 4.3	Processing Layer Internal Architecture	80
Figure 4.4	Instructor Dashboard UI Wireframe	85
Figure 4.5	Alert Notification Workflow Diagram	89
Figure 4.6	Docker/Kubernetes Deployment Architecture	93
Figure 5.1	Data Collection Polling Architecture	100
Figure 5.2	Feature Engineering Pipeline — Processing Stages	109
Figure 5.3	Feature Correlation Heatmap — Training Dataset	114
Figure 5.4	Machine Learning Pipeline Architecture Diagram	117
Figure 5.5	Ensemble Weight Optimization — Grid Search Results	124
Figure 5.6	REST API Endpoint Map	128
Figure 5.7	Instructor Dashboard — Main Risk Overview Panel	132
Figure 5.8	Student Detail Panel — Feature Contribution Chart	134
Figure 6.1	Entity-Relationship Diagram — PostgreSQL Schema	140
Figure 6.2	Database Partitioning Strategy — Interactions Table	150
Figure 8.1	Dataset Enrollment Distribution by Course Category	176
Figure 8.2	Class Distribution — Dropout vs non-dropout	177
Figure 8.3	Model Performance Comparison — All Classifiers.....	179
Figure 8.4	Relative Performance Radar Chart — Five Metrics	180
Figure 8.5	Confusion Matrix — Ensemble Model	183
Figure 8.6	ROC Curves — All Models Compared	184
Figure 8.7	Precision-Recall Curves — All Models	184
Figure 8.8	Temporal Accuracy Evolution — Weeks 1 through 12	186
Figure 8.9	Early Warning Window Analysis — Week 3 vs Week 6	188
Figure 8.10	Feature Importance Bar Chart — Top 15 Features.....	190
Figure 8.11	Feature Importance by Category — Grouped Bar Chart	191
Figure 8.12	Dropout Rate by Course Category — Comparison	194

Figure 8.13 Weekly Dropout Distribution — All Three Courses	195
Figure 8.14 Cumulative Dropout Curve — 12-Week Timeline	196
Figure 8.15 UAT Task Completion Time — Five Instructors	199

List of table

Table No.	Caption	Page No.
Table 1.1	Research Objectives Summary Matrix	17
Table 2.1	Summary of Prior Dropout Prediction Studies (2013–2023)	34
Table 2.2	Machine Learning Algorithms Applied in Prior Research — Pros and Cons	38
Table 2.3	Feature Categories Used Across Major Prior Studies	44
Table 2.4	Gap Analysis — Limitations of Prior Research	51
Table 3.1	Technical Stack Feasibility Assessment	56
Table 3.2	Operational Feasibility — Stakeholder Impact Matrix.....	60
Table 3.3	Economic Feasibility — Cost-Benefit Analysis	64
Table 3.4	Schedule Feasibility — Development Milestones.....	66
Table 3.5	Overall Feasibility Summary Matrix	67
Table 4.1	System Modules and Layer Assignment	73
Table 5.1	Technology Stack — Components and Justifications.....	96
Table 5.2	Interaction Event Categories and Fields Captured.....	101
Table 5.3	Data Cleaning Rules Applied to Raw Events.....	105
Table 5.4	Complete Feature Engineering Variables (All 27 Features)	111
Table 5.5	Feature Importance Scores — Random Forest Permutation Importance	115
Table 5.6	Hyperparameter Configurations — All Three Classifiers	122
Table 5.7	REST API Endpoint Definitions	129
Table 6.1	Students Table Schema	142
Table 6.2	Courses Table Schema	143
Table 6.3	Enrollments Table Schema	144
Table 6.4	Interactions Table Schema	145
Table 6.5	Predictions Table Schema	146
Table 6.6	Alerts Table Schema	147
Table 6.7	Database Index Strategy	149
Table 7.1	Unit Test Summary by Module	156
Table 7.2	Integration Test Cases and Results	160
Table 7.3	Load Test Performance Results	165
Table 7.4	UAT Task Completion Results	168
Table 7.5	Security Test Cases and Results	172
Table 8.1	Dataset Statistics — Enrollment, Dropout, and Completion by Course	175
Table 8.2	Model Performance Comparison — All Metrics	178
Table 8.3	Confusion Matrix Values — Ensemble Model	182
Table 8.4	TemporaAccuracy Course Week — Ensemble Model	185

Table 8.5	Feature Importance Rankings — Top 20 of 27	189
Table 8.6	Dropout Rate and Peak Dropout Week by Course Category	193
Table 8.7	Dropout Timing Distribution — Weekly and Cumulative	195
Table 8.8	UAT Ease Ratings and Completion Times	198
Table 8.9	Comparative Performance — This Study vs Prior Research.....	201
Table A1	Random Forest Full Hyperparameter Tuning Grid	232
Table A2	Gradient Boosting Full Hyperparameter Tuning Grid	233
Table A3	Deep Neural Network Architecture Search Results	234
Table A4	SMOTE Oversampling Impact on Training Set	235

List of Abbreviations

Abbreviation Full Form

AI	Artificial Intelligence
AUC	Area Under the ROC Curve
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CPU	Central Processing Unit
CRUD	Create, Read, Update, Delete
DNN	Deep Neural Network
EDM	Educational Data Mining
GB	Gradient Boosting
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HOD	Head of Department
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
JSON	JavaScript Object Notation
JWT	JSON Web Token
LA	Learning Analytics
LMS	Learning Management System
LSTM	Long Short-Term Memory
MOOC	Massive Open Online Course
ML	Machine Learning
MUI	Material-UI (React component library)
NLP	Natural Language Processing
PDPA	Personal Data Protection Act
RF	Random Forest
REST	Representational State Transfer
ROC	Receiver Operating Characteristic
SDT	Self-Determination Theory
SMOTE	Synthetic Minority Oversampling Technique
SQL	Structured Query Language
SVM	Support Vector Machine

TLS	Transport Layer Security
TTL	Time-To-Live
UAT	User Acceptance Testing
UI	User Interface
UTC	Coordinated Universal Time
WAL	Write-Ahead Logging