



Customer Risk Profiling System

G.A.R. GUNASEKARA
(Reg. No.: MS24019590)

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 SYSTEM)

December 2025

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. Kanishka Yapa

Approved for MSc. Research Project:

MSc in IT Program Co-Ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

I hereby declare that this thesis, entitled “Customer Risk Profiling System for the Insurance Industry”, is the result of my own independent research and work, except where specific references are made and duly acknowledged. This work has not been submitted in whole or in part, in any form, to the Sri Lanka Institute of Information Technology or to any other institution of higher education for the purpose of academic assessment, qualification, or award. All sources of information, data, and ideas have been properly cited and acknowledged according to academic standards.

Sign: 

Student Name: G.A.R.Gunasekara

Date: 2025-11-03

ABSTRACT

Development of a Customer Risk Profiling Framework for the Insurance Industry Using Machine Learning and Centralized Insurance GRIB

Candidate: Radeena Gunasekara

Program: MSc in Information Systems

Supervisor: [Mr. Kanishka Yapa]

Date: December 2025

This thesis addresses the critical need for advanced, ethical risk quantification in the motor insurance sector, currently hampered by fragmented data and limited cross-company fraud visibility. The primary objective was to design and validate a **Customer Risk Profiling System (CRPS)** that integrates heterogeneous data sources and utilizes Machine Learning (ML) for dynamic risk scoring. The methodology involved aggregating data streams including claims, premiums, policy history, and external **PEP/AML compliance scores** and employing **Gradient Boosted Trees (GBTs)** to achieve a high classification **Accuracy of \$0.7561\$** and a **ROC-AUC of \$0.8982\$**. Empirical findings confirmed the predictive power of **behavioral features** over linear demographic metrics, validating the choice of non-linear ensemble models. The CRPS successfully segments customers into Low, Medium, and High-Risk tiers, enabling targeted intervention. Crucially, the system embeds **Explainable AI (XAI)** using SHAP values and a continuous **Feedback Loop** to maintain accuracy against concept drift, ensuring **auditability and ethical governance** against potential bias. The study concludes by proposing the **Insurance National Grid (NIG)**, a centralized platform designed to connect all insurers to the regulator. The NIG would enforce **data standardization** and enable **cross-company fraud detection**, magnifying the CRPS's impact from a firm-specific tool to a national strategic asset, thereby promoting market efficiency, compliance, and sustained sector resilience.

Keywords: Machine Learning, Risk Profiling, Ensemble Methods, Explainable AI (XAI), Fraud Detection, Data Governance, Insurance National Grid (NIG), Motor Insurance, Regulatory Compliance.

Acknowledgement

The successful completion of this research work represents a crucial milestone in my academic career, and it would not have been possible but for the encouragement, advice, and support of numerous individuals and institutions that contributed in various manners. I am deeply indebted to all those persons who remained consistent throughout this process, and it is with sincere gratitude that I express my thanks here. Most importantly, I wish to extend my deepest gratitude to my supervisor, Mr. Kanishka Yapa, whose consistent guidance, patience, and constructive criticism were instrumental in determining the direction of this research. Throughout the process, starting from the early stages of conceptualizing the research problem until the thesis was completed, his critical insight and sharp eye for detail ensured academic rigor and methodological efficiency. His ability to question my ideas and yet encourage me to think more innovatively in my work has been a savior. I am particularly grateful for the time he took to examine my work and for his guidance during those times when the research seemed gigantic. I am also thankful to Mr. Chinthana Samarasekera, a practitioner in business whose business knowledge and experience were the key bridge between research study and practice. His comments helped me to value the bigger picture of the insurance sector, namely customer risk profiling and fraud identification. Through case studies and industry practice, he enabled me to place my research in a practical context so that the findings would be meaningful and helpful. His capacity to offer industry insights and guidance greatly enhanced the value of this project.

Particular gratitude is due to Mr. Ravindra Rathnayaka, HNB General and HNB Assurance IT Head, whose inspiration and institutional support opened the doors for me to learn about the operational and technological challenges faced by insurance companies. He provided not just exposure to hard realities but also motivation by indicating the direction of filling the gap between organizational needs and technological innovation. His contribution was enormous towards making me understand the potential for digital transformation to redirect the insurance industry. I also appreciate the contribution of Mr. Nuwan Athukorala a data analytics expert whose technical expertise helped me construct the analytical framework of this research. His recommendations on advanced modeling techniques, feature engineering, and data unification helped me enhance the technical basis of the system used in this study.

His advice on analytics not only improved the accuracy of my models but also enhanced my understanding on how data science can be applied ethically in sensitive industries such as insurance. Apart from professional and academic help, I owe the most to my family members. Their support, patience, and ongoing inspiration have been the bulwarks of this research. They encouraged me during long nights of studying, boosted me during hardships, and celebrated each small triumph along the way. Without their moral support and belief in my abilities, it would have been much more difficult to complete this work. Lastly, this thesis is a testimony to the combined efforts of mentors, industry leaders, and family members who have all, in their own ways, ensured that this is done. I am still deeply indebted to all of them and hope that this work is a testimony to the tremendous guidance they have offered.

Table of Contents

DECLARATION	3
ABSTRACT	4
Acknowledgement	5
List of Figures	12
List of Tables	12
Chapter 1: Introduction	13
1.1 Background	13
1.2 Problem Statement	15
1.3 Research Objectives	16
1.4 Research Questions	17
1.5 Scope and Limitations	18
Chapter 2: Literature Review	20
2.1 Introduction to the Modern Paradigm of Customer Risk Profiling	20
2.1.1 The Inadequacy of Traditional Underwriting Architectures	20
2.1.2 The Digital Transformation Catalyst	21
2.1.3 The Value Proposition of Multi-Faceted Data Integration	22
2.1.4 The Centrality of Behavioral, Contextual, and Compliance Factors	23
2.2 Domain Research Review: Three Pillars of Modern Risk Profiling	24
2.2.1 Machine Learning and Ensemble Methods for Predictive Modeling	24
2.2.2 Data Integration, Feature Engineering, and Anomaly Detection	25
2.2.3 Ethical AI, Explainability, and Regulatory Compliance	26
2.3 Machine Learning Methods, Data Integration, and Advanced Feature Engineering in Risk Profiling	27
2.4 Ethical and Regulatory Imperatives, Governance, and Research Gap Synthesis	31

2.5 Gaps in Existing Systems and Research	36
Chapter 3: System Design and Methodology	39
3.1 Research Design and Data Description	39
3.2 System Context and Modules.....	43
3.2.1 Data Aggregation Module: Resolving the Fragmentation Imperative	43
3.2.2 Feature Engineering Module: The Creation of Predictive Intelligence.....	44
3.2.3 Risk Scoring Model Module: Precision, Interpretability, and Validation	45
3.2.4 Decision System Module: Translating Score to Action	46
3.2.5 Feedback Loop Module: Enforcing Continuous Relevance	47
3.3 Data Preprocessing and Feature Engineering.....	48
3.3.1 Data Preprocessing	49
3.3.2 Feature Engineering.....	50
3.4.1 Risk Scoring Model Development: Maximizing Accuracy and Interpretability	51
3.4.2 Customer Tracking System: Operationalizing Predictive Intelligence	53
3.4.3 Conclusion of System Context and Modules	55
3.5 Feedback Loop and Ethical Governance.....	55
3.5.1 The Continuous Feedback Loop: Enforcing Dynamic Relevance	55
3.5.2 Ethical Concerns, Compliance, and Accountability Governance.....	58
3.6 Conceptual Diagram: Customer Risk Profiling System (CRPS)	60
3.7 Class Diagram	63
Chapter 4: Results and Findings	65
4.1 Introduction	65
4.2 Data Overview.....	65
4.3 Feature Engineering	67
4.4 Exploratory Data Analysis (EDA)	68

4.5 Multi-Model Training and Evaluation Results.....	76
4.6 System Implementation and UI Findings.....	80
4.7 Multi-Source Data Collection	84
4.8 Synthesis of Findings and Analytical Facts	93
Chapter 5: Discussion and Challenges.....	96
5.1 System Implementation and Operational Challenges	96
5.1.1 Data Integration and Pipeline Complexity	96
5.1.2 Model Interpretability and Underwriter Adoption	97
5.1.3 Scalability and Real-Time Performance Requirements.....	97
5.2 Data Limitations and Mitigation Strategies	98
5.2.1 Incomplete, Sparse, and Imbalanced Data.....	98
5.2.2 Historical Bias in Training Data.....	98
5.2.3 External Data Dependencies and Uncertainty	99
5.3 Algorithmic Limitations and Performance Management.....	99
5.3.1 Risk of Overfitting and Generalization	99
5.3.2 Accuracy vs. Interpretability Trade-off.....	100
5.3.3 Data Drift and Concept Drift Vulnerability.....	100
5.4 Regulatory and Ethical Challenges	101
5.4.1 Data Privacy and PII Compliance	101
5.4.2 Bias Mitigation and Algorithmic Fairness.....	101
5.4.3 Auditability and Accountability	101
5.5 Strategic Implications of Overcoming Challenges.....	102
Chapter 6: Proposal of a National Insurance Grid for Sri Lanka.....	103
6.1 Introduction: The Imperative for Centralized Intelligence	103
6.2 Motivation and Rationale: Addressing Systemic Inefficiencies	104

6.2.1 Challenges in the Current Insurance Landscape.....	104
6.2.2 National Insurance Grid Objectives: A Vision for Unified Governance.....	105
6.3 Conceptual Overview of the National Insurance Grid	106
6.3.1 Insurer Portals: The Decentralized Gateway to Centralization	106
6.3.2 Central Analytics Platform (CAP): The Predictive Core	107
6.3.3 Regulatory Dashboards: Empowering the IRCSL.....	108
6.4 Technical Architecture and Data Processing	109
6.4.1 Data Preprocessing for Cross-Company Consistency	110
6.4.2 Feature Engineering for Unified Risk Prediction	111
6.5 Machine Learning Risk Scoring and Segmentation.....	111
6.6 Portfolio-Level Insights and Strategic Value	112
6.7 Implementation Strategy, Governance, and Benefits.....	113
6.7.1 Phased Implementation Approach.....	113
6.7.2 Ethical and Legal Governance.....	114
6.7.3 Expected Benefits: A National Advantage	115
6.8 Innovation, Originality, and Future Extensions	115
Chapter 7: Conclusion, Practical Implications, and Future Research.....	117
7.1 Summary of Findings: The Validation of Predictive Governance	117
7.2 Practical Implications for the Insurance Industry	119
7.2.1 Operational Efficiency and Underwriting Automation	119
7.2.2 Advanced Fraud Detection and Systemic Mitigation.....	119
7.2.3 Customer Segmentation, Personalization, and Retention Strategy	120
7.2.4 Portfolio-Level Risk Management and Strategic Decision-Making	120
7.2.5 Enhanced Regulatory Compliance and Systemic Trust	121
7.3 Recommendations for Future Research	121

7.3.1 Integration of Advanced External Data Sources and Privacy Preservation	121
7.3.2 Cross-Product Risk Modeling and Cross-Selling Optimization.....	122
7.3.3 Advanced Modeling Techniques for Dynamic Resilience	122
7.3.4 Enhanced Fraud Detection with Network Analytics and Explainability.....	123
7.3.5 User Experience, Decision Support, and Ethical Governance	123
Chapter 8: Conclusion, Contributions, and Future Research.....	124
8.1 Summary of Core Research Contributions.....	124
8.1.1 Machine Learning-Driven Risk Profiling Architecture.....	124
8.1.2 Compliance with Ethical and Regulatory Imperatives	125
8.1.3 Continuous Model Retraining and Feedback Loop Dynamism	126
8.1.4 Insurance NIG: An Industry Centralized Platform Innovation	126
8.2 Practical Implications for the Insurance Industry	127
8.2.1 Operational Efficiency and Strategic Resource Allocation.....	127
8.2.2 Fraud Prevention, Detection, and Mitigation	128
8.2.3 Customer-Centric Management and Market Trust.....	128
8.2.4 Portfolio-Level Strategic Guidance	128
8.3 Recommendations for Future Research	129
8.3.1 Integration of Advanced External Data and Secure Linkage	129
8.3.2 Next-Generation Modeling and Adaptive Resilience.....	129
8.3.3 Real-Time Decision Support and UX Optimization.....	130
8.3.4 Ethical AI, Governance, and Multi-Line Extension	130
References.....	131

List of Figures

Figure 1 : Conceptual Framework Diagram.....	61
Figure 2 : Conceptual Framework Diagram.....	62
Figure 3 : Class Diagram.....	63
Figure 4 : Data Cleaning	66
Figure 5 : Correlation Analysis	68
Figure 6 : Correlation with Customer Risk Level.....	71
Figure 7 : Gender-wise Distribution Analysis for Risk Profiling	72
Figure 8 : Distribution Analysis of Core Predictive Features.....	73
Figure 9 : Density Plot of Sanction Score by Risk Level	74
Figure 10 : Pairwise Relationships and Risk Classification	75
Figure 11 : Confusion Matrix	77
Figure 12 : Confusion Matrix	78
Figure 13 : Model Accuracy Comparison	78
Figure 14 : System UI.....	80
Figure 15 : System UI.....	81
Figure 16 : System UI.....	81
Figure 17 : Sample of Survey Results.....	86
Figure 18 : Sample of Survey Results.....	87
Figure 19 : Sample of Survey Results.....	88
Figure 20 ; Sample of Survey Results.....	89
Figure 21 : Sample Survey Results	90
Figure 22 ; Sample of Survey Results.....	91
Figure 23 : HNBGI Core system Dataset	92
Figure 24 : Requirements Gathering Form	92

List of Tables

Table 1 : Gaps in Existing Systems and Research.....	38
Table 2 : Dataset Summary.....	65
Table 3 : Derived Features	68
Table 4 : Model Comparison Summary.....	80
Table 5 : The risk scores	112