



Table of Contents

Title Page	i
Table of Contents	iv
Table of Contents - Tables	vii
Table of Contents – Figures	ix
Table of Contents– Appendices	xii
List of Abbreviations.....	xiii
<i>Intisari</i>	xiv
Abstact.....	xv
CHAPTER I. INTRODUCTION.....	1
1.1 Research Background.....	1
1.2 Problem Statement	4
1.3 Research Questions	5
1.4 Research Objectives	5
1.5 Research Motivation	6
1.6 Research Contributions	7
1.7 Thesis Structure.....	8
CHAPTER II. THEORETICAL FRAMEWORK	11
2.1 Fraud.....	11
2.2 Fraud Detection in Accounting Information System	13
2.3 Expenditure Cycle in Accounting Information System	14
2.3.1 Purchasing Procedures	16
2.3.2 Cash Disbursements Procedures	19
2.4 Challenges in Fraud Detection	22
2.4.1 Imbalance of Fraudulent vs. Legitimate Transactions	22
2.4.2 Data Manipulation by Fraudsters	23
2.4.3 Limitations of Traditional Rule-Based Controls.....	24
2.4.4 Evolving Nature of Fraudulent Activities	24
2.4.5 Cost of False Positives	25
2.5 Machine Learning in Fraud Detection	26
2.5.1 Logistic Regression.....	27
2.5.2 Random Forest	28
2.5.3 XGBoost (Extreme Gradient Boosting)	28
2.5.4 Support Vector Machines (SVM)	29
2.5.5 K-Nearest Neighbors (KNN)	29
2.5.6 Convolutional Neural Networks.....	30
2.6 Evaluation Metrics for Machine Learning Models	31
2.6.1 Accuracy.....	31
2.6.2 Precision, Recall, and F1-Score	32
2.6.3 Receiver Operating Characteristic (ROC) Curve & AUC Score	32



2.6.4 Confusion Matrix	34
2.6.5 Cost-Sensitive Metrics in Fraud Detection	34
CHAPTER III. METHODOLOGY	35
3.1 Research Framework.....	35
3.2 Dataset.....	35
3.3 Data Preprocessing.....	36
3.4 Machine Learning Models	37
3.4.1 Logistic Regression.....	37
3.4.2 Random Forest	37
3.4.3 XGBoost.....	38
3.4.4 Support Vector Machine (SVM).....	38
3.4.5 K-Nearest Neighbors (KNN)	38
3.4.6 Neural Networks	38
3.5 Model Evaluation	39
3.5.1 Accuracy.....	39
3.5.2 Precision	39
3.5.3 Recall.....	39
3.5.4 F1-Score	40
3.5.5 Area Under the ROC Curve (AUC-ROC).....	40
3.6 Hyperparameter Tuning	40
3.7 Research Procedure	40
3.8 Research Framework.....	41
CHAPTER IV. DATA ANALYSIS MODEL DEVELOPMENT, AND FRAMEWORK INTEGRATION	44
4.1 Descriptive Statistics Analysis Results	44
4.1.1 Dataset Overview	44
4.1.2 Variable Types and Structure.....	46
4.1.3 Summary Statistics.....	47
4.2 Exploratory Data Analysis (EDA)	51
4.2.1 Fraud Class Distribution	51
4.2.2 Feature Distributions	53
4.3.3 Correlation Analysis	57
4.3.4 Transaction Type Analysis.....	61
4.3.5 Zero Balance Indicators	63
4.4 Data Preprocessing	65
4.4.1 Handling Categorical Variables	65
4.4.2 Log-Transforming Skewed Features.....	66
4.4.3 Zero Balance Indicators	66
4.4.4 Scaling Numerical Features	66
4.4.5 Preparing Final Feature Set.....	67
4.4.6 Handling Class Imbalance.....	67
4.5 Model Building	69
4.5.1 Logistic Regression.....	70



4.5.2 Random Forest	76
4.5.3 XGBoost.....	83
4.5.4 Support Vector Machines (SVM)	90
4.5.5 K-Nearest Neighbors (KNN)	96
4.5.6 Neural Networks	103
4.6 Model Evaluation Summary	110
4.7 Fraud Indicator Analysis.....	117
4.7.1 Key Predictors of Fraudulent Transactions.....	117
4.7.2 Alignment with Expenditure Cycle Fraud Patterns	119
4.8 Proposed Integration Framework in Accounting Information System (AIS)	119
4.8.1 Framework Components	120
4.8.2 Implementation Context.....	122
CHAPTER V. CONCLUSION, IMPLICATION AND CONSTRAINTS...	123
5.1 Conclusion	123
5.2 Implications.....	125
5.3 Constraints	126
References	128