

# TABLE OF CONTENTS

<b>COVER</b>	
<b>ORIGINALITY STATEMENT OF THE THESIS</b>	
<b>THESIS SUPERVISOR’S APPROVAL</b>	
<b>THESIS EXAMINATION COMMITTEE</b>	
<b>ABSTRACT .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>vii</b>
<b>TABLE OF CONTENTS.....</b>	<b>ix</b>
<b>LIST OF FIGURES .....</b>	<b>xi</b>
<b>LIST OF TABLES .....</b>	<b>xii</b>
<b>CHAPTER I INTRODUCTION.....</b>	<b>1</b>
1.1. Motivation .....	1
1.2. Problem Formulation.....	5
1.3. Scopes of Problem.....	5
1.4. Research Purpose .....	6
1.5. Research Methodology.....	6
<b>CHAPTER II REVIEW OF LITERATURE .....</b>	<b>8</b>
2.1. Customer Lifetime Value .....	8
2.2. Retention Model for CLV Measurement .....	9
2.3. Data Cleansing .....	10
2.4. Data Transformation .....	11
2.5. Data Masking .....	12
2.6. Random Forest Regression.....	14
2.7. XGBoost (Extreme Gradient Boosting).....	16
2.8. Evaluation Metrics and Performance Comparison.....	19
<b>CHAPTER III RESEARCH METHODOLOGY .....</b>	<b>23</b>
3.1. Data Collection.....	23
3.2. Explanatory Data Analysis.....	24
3.3. Data Cleansing .....	25

3.4. Modelling .....	27
3.5. Evaluation.....	27
3.6. Analysis.....	28
<b>CHAPTER IV RESULTS AND ANALYSIS.....</b>	<b>30</b>
4.1. Model Performance Comparison.....	31
4.2. Residual Analysis.....	34
4.2.1. Predicted Value vs. Actual Value .....	34
4.2.2. Residual Value vs. Predicted Value.....	36
4.3. Feature Importance Analysis.....	37
4.4. Research Journal .....	40
<b>CHAPTER V CONCLUSION .....</b>	<b>58</b>
<b>REFERENCES</b>	
<b>APPENDIX</b>	

## LIST OF FIGURES

Figure 1 - Research Methodology .....	23
Figure 2 - Performance Metrics RMSE and MAE Comparison .....	32
Figure 3 - Performance Metrics R-Squared Comparison.....	33
Figure 4 - Residual Plot Predicted vs. Actual Value.....	34
Figure 5 - Residual Plot Residual vs. Predicted Value .....	35
Figure 6 - Feature Importance Random Forest .....	37
Figure 7 - Feature Importance XGBoost.....	39

## **LIST OF TABLES**

Table 1 Machine Learning Models Advantages and Disadvantages .....	2
Table 2 - Model Performance Comparison .....	30