

UNIVERSITI TEKNOLOGI MARA

**PREDICTING CUSTOMER CHURN IN
TELECOMMUNICATION SERVICE
PROVIDER INDUSTRY USING
RANDOM FOREST**

**WAN MUHAMMAD NAQIB ZAFRAN BIN WAN
ROSLAN**

**BACHELOR OF COMPUTER SCIENCE AND
MATHEMATICS (Hons.)**

JULY 2023

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to my supervisor, Ahmad Nazmi Bin Fadzal for guiding me for two semesters. I would also like to express my appreciation to my family and friends, who support and encourage me to complete this thesis throughout my study. Thank you to all the individuals who actively participated in my thesis. As I concluded this thesis, it has been a rewarding journey as I reflected on the path I have taken. Completing this thesis has been filled with challenges yet I have learned to overcome it.

ABSTRACT

This project addresses the challenge of customer churn in the Telecommunications Service Provider (TSP) industry by focusing on the Random Forest algorithm for predictive modeling. This study aims to thoroughly explore the Random Forest algorithm, develop a robust Random Forest customer churn predictive model, and evaluate its performance in predicting customer churn within the internet service provider sector. The specific objectives include studying the complexity of the Random Forest algorithm, constructing a model adjusted to accurately predict customer churn, and conducting thorough testing and evaluation of the model's accuracy. Through many experimentation, it was found that a model with 20 trees, a maximum depth of 5, and a maximum of 8 features yielded the highest accuracy at 79%, with an area under the curve of 0.79 for the Receiver Operating Characteristics. The outcomes of this research are poised to contribute significantly to the improvement of revenue, customer satisfaction, and provide valuable insights for data scientists and analysts engaged in similar predictive modeling endeavors within the TSP industry.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	II
STUDENT DECLARATION	III
ACKNOWLEDGEMENT	IV
ABSTRACT	V
LIST OF FIGURES	IX
LIST OF TABLES	X
 CHAPTER ONE: INTRODUCTION	
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Objective	3
1.4 Scope	4
1.5 Significant of Study	5
1.6 Overview of Research Framework	6
1.7 Gantt Chart for Project	7
1.8 Conclusion	8
 CHAPTER 2: LITERATURE REVIEW	
2.1 Predictive analytic	9
2.1.1 Application of predictive analytic	10
2.2 Random Forest algorithm	11
2.2.1 Overview of the Random Forest	11
2.2.2 Steps for Random Forest	11
2.2.3 Application of Random Forest	13
2.2.4 Related work using Random Forest	14
2.3 Telecommunication service provider industry	18
2.3.1 Overview of customer churn in the telecommunication service provider	18

2.3.2	Issues in telecommunication service provider industry	18
2.4	Similar application	19
2.5	Implication of literature review	23
2.6	Conclusion	24

CHAPTER 3: METHODOLOGY

3.1	Research Framework	25
3.2	Preliminary Study	29
3.2.1	Data collection	29
3.2.2	Data description	29
3.2.3	Data preprocessing	31
3.3	Design and implementation phase	32
3.3.1	System Architecture	32
3.3.2	System workflow	34
3.3.3	User interface design	36
3.3.4	Prototype Implementation	37
3.4	Performance evaluation	38
3.4.1	Confusion matrix	38
3.4.2	Receiver Operating Characteristics Curve	39
3.5	Conclusion	40

CHAPTER 4: RESULT AND FINDING

4.1	Conceptual Framework	41
4.2	Program Codes for Random Forest Algorithm	44
4.2.1	Decision Tree	44
4.2.2	Random Forest	50
4.2.3	Training Model	54
4.3	Prototype Interface	56
4.4	Evaluation Results	57
4.4.1	Confusion Matrix	57
4.4.2	Receiver Operating Characteristics Curve	59
4.4.3	Hyperparameter Tuning	60
4.5	Conclusion	62