

UNIVERSITI TEKNOLOGI MARA

**PREDICTIVE MODEL FOR BANK CUSTOMER
CHURN USING RANDOM FOREST**

NUR IFFAH MALIAH BINTI ROZI

**BACHELOR OF COMPUTER SCIENCE (HONS.)
COLLEGE OF COMPUTER INFORMATICS AND MATHEMATICS**

FEBRUARY 2024

ACKNOWLEDGEMENT

All praise and thanks to Allah SWT and His great mercy for allowing me to finish this Final Year Project. I extend my deepest appreciation to all individuals and entities whose collective contributions have played a pivotal role in the successful completion of this project, which delves into the creation and application of a predictive model for bank customer churn employing the Random Forest algorithm. Foremost, I express my profound gratitude to my supervisor, Madam Sharifah Nurulhikmah binti Syed Yasin for her exceptional guidance, steadfast support, and invaluable insights that have significantly shaped this research.

My sincere thanks are extended to Madam Ummu Fatimah binti Mohd Bahrin for her exceptional consideration, expertise, and support. Moreover, I am indebted to my family and friends for their enduring encouragement and understanding throughout this academic pursuit. Their unwavering support has been a constant source of motivation.

I would be remiss not to acknowledge the wider research community and my peers for their meaningful contributions. Engaging in insightful discussions and benefiting from shared knowledge have undeniably enriched the quality and robustness of the developed predictive model. Lastly, I express my gratitude to all those unmentioned but integral to the completion of this work. This thesis represents a collaborative effort, and I am profoundly thankful for the collective support and collaboration that has shaped its success.

ABSTRACT

This project delves into the design and implementation of a predictive model aimed at forecasting bank customer churn, employing the Random Forest algorithm. In the dynamic landscape of the banking industry, the phenomenon of customer churn presents a formidable challenge. This study seeks to confront and mitigate this challenge by harnessing the efficacy of Random Forest, a machine learning ensemble technique renowned for its resilience and precision. To achieve this, the research involves the meticulous collection of diverse customer data, encompassing transaction histories, demographic details, and various customer interactions. The data is utilized to train and validate the Random Forest model, with a comprehensive evaluation of its predictive performance metrics such as accuracy, precision, recall, and F1-score. Notably, the Random Forest model achieved an accuracy score of 79.97%. The anticipated outcomes of this research not only contribute to the progression of predictive modeling within the banking sector but also furnish financial institutions with actionable insights to proactively discern and address customer churn.

TABLE OF CONTENTS

| CONTENT | PAGE |
|-------------------------------------------------------|------------|
| SUPERVISOR APPROVAL..... | ii |
| STUDENT DECLARATION..... | iii |
| ACKNOWLEDGEMENT..... | iv |
| ABSTRACT..... | v |
| LIST OF FIGURES..... | ix |
| LIST OF TABLES..... | x |
| | |
| CHAPTER 1: INTRODUCTION..... | 1 |
| 1.1 Background of Study..... | 1 |
| 1.2 Problem Statement..... | 2 |
| 1.3 Objective..... | 3 |
| 1.4 Project Scope..... | 4 |
| 1.5 Project Significance..... | 5 |
| 1.6 Overview of Research Framework..... | 6 |
| 1.7 Conclusion..... | 8 |
| | |
| CHAPTER 2: LITERATURE REVIEW..... | 9 |
| 2.1 Overview of Bank Customer Churn..... | 9 |
| 2.1.1 Causes of Bank Customer Churn..... | 9 |
| 2.1.2 How Bank Customer Churn Affects the Market..... | 10 |
| 2.1.3 Bank Customer Churn Detection Strategies..... | 11 |
| 2.2 Bank Customer Churn Detection Algorithm..... | 13 |
| 2.2.1 Overview of Algorithm..... | 13 |
| 2.2.2 How Does Random Forest Algorithm Work..... | 16 |
| 2.2.3 Random Forest Algorithm Feature..... | 17 |

| | | |
|-------|-------------------------------------------------------|----|
| 2.2.4 | Advantage of Random Forest Algorithm | 19 |
| 2.3 | Similar Application | 20 |
| 2.3.1 | Implementation of Algorithm in Various Problems | 32 |
| 2.4 | Implication of Literature Review | 45 |
| 2.5 | Conclusion | 46 |

CHAPTER 3: METHODOLOGY 47

| | | |
|-------|--------------------------------------------------|----|
| 3.1 | Overview of Research Methodology Framework | 47 |
| 3.2 | Preliminary Study | 52 |
| 3.2.1 | Knowledge Acquisition..... | 53 |
| 3.3 | Data Collection | 54 |
| 3.3.1 | Data Description | 56 |
| 3.3.2 | Data Pre-Processing..... | 56 |
| 3.4 | Design Phase..... | 58 |
| 3.4.1 | System Architecture..... | 60 |
| 3.4.2 | Flowchart..... | 61 |
| 3.4.3 | Pseudocode | 62 |
| 3.4.4 | Interface Design..... | 63 |
| 3.5 | System Development | 64 |
| 3.5.1 | Software Recommendation | 64 |
| 3.5.2 | Hardware Recommendation..... | 65 |
| 3.6 | Evaluation Phase..... | 66 |
| 3.6.1 | Accuracy of Algorithm | 67 |
| 3.7 | Documentation..... | 69 |
| 3.8 | Gantt Chart..... | 70 |
| 3.9 | Conclusion | 71 |

CHAPTER FOUR: RESULT AND FINDING 72

| | | |
|-------|-------------------------------------------|----|
| 4.1 | Algorithm Application..... | 72 |
| 4.2 | Explanation of the System Framework | 73 |
| 4.2.1 | Dataset Preprocessing..... | 73 |