

Universiti Teknologi MARA

**AIRLINE FLIGHT DELAY
PREDICTION USING NAÏVE BAYES
ALGORITHM**

AHMAD ADIB BAIHAQI BIN SHUKRI

**Thesis submitted in fulfilment of the requirements
for Bachelor of Computer Science (Hons.) College
of Computing, Informatics and Mathematics**

JANUARY 2024

ACKNOWLEDGEMENT

Alhamdulillah, praises and thank to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks goes to my supervisor, Ts. Dr Syarifah Adilah Binti Mohamed Yusoff, for her invaluable guidance, patience, and continuous encouragement. Her expertise and insights have been instrumental in shaping the direction of my project. I would like to express my gratitude to Madam Ummu Fatimah Binti Mohd Bahrin, who served as my CSP600 and CSP650 lecturer. Without her guidance and support, I might be loss through the entire semester.

Special appreciation also goes to my beloved parents and family. Their unwavering support, encouragement, and understanding have been the foundation of my success. I overcame obstacles and succeeded in reaching the end line because of their encouragement and belief in my abilities. They contributed greatly to the development of this project by their presence and support in difficult times. I've really been grateful to them all this time for their love, patience and understanding.

Last but not least, I would like to give my gratitude to my friends my dear friends for their unwavering support and encouragement throughout the completion of this project. They have served as a consistent source of inspiration and motivation. Their assistance has been priceless, from listening when I was frustrated to providing insightful commentary and suggestions.

ABSTRACT

The aviation industry plays a critical role in global transportation, facilitating economic growth and revolutionizing travel. However, flight delays have become a growing concern, impacting both airlines and passengers. This study aims to study the Naïve Bayes algorithm for flight delay prediction. The objective is to develop a reliable flight delay prediction model using the Naïve Bayes algorithm and evaluate its performance. The data set that records flight delay and cancellation data from U.S Department of Transportation's (DOT) was used for the prediction. Three algorithms (Gaussian Naïve Bayes, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM)) were trained and tested to complete the binary classification of flight delays. Parameter tuning also done on Gaussian Naïve Bayes by changing its parameter. The evaluation of algorithms was fulfilled by comparing the values of accuracy, specificity and ROC AUC score. These measures were weighted to adjust the imbalance of the selected data set. The comparative analysis showed that the Gaussian Naïve Bayes has the best performance with an accuracy of 93% and KNN has the worst performance with ROC AUC score 63%. The Naïve Bayes classifier generally have better performance over other base classifiers.

Keywords: aviation industry, flight delays, Naïve Bayes algorithm, prediction model, machine learning, U.S Department of Transportation (DOT), data set, binary classification, Gaussian Naïve Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), parameter tuning, accuracy, specificity, ROC AUC score, imbalance, comparative analysis, classifier performance.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Research Questions	4
1.4 Objectives	4
1.4 Project Scope	4
1.5 Project Significance	5
1.6 Overview of Research Framework	6
1.7 Conclusion	7
CHAPTER TWO: LITERATURE REVIEW	8
2.1 Airline Industry	8
2.1.1 Current Operations	9
2.1.2 Problem with current operations	10
2.1.3 Importance of Artificial Intelligence to handle this issue	11
2.2 Naïve Bayes Algorithm	12
2.2.1 General Concept	12

2.2.2	Advantages and Disadvantages	14
2.2.3	Naïve Bayes Classification Architecture	15
2.3	Implementation of Naïve Bayes in various problems	16
2.4	Similar works	23
2.5	Implication of Literature Review	25
2.6	Conclusion	26
CHAPTER THREE: METHODOLOGY		27
3.1	Overview of Research Methodology	27
3.1.1	Detailed Research Framework	27
3.2	Preliminary Phase	34
3.2.1	Literature Review	34
3.2.2	Data Collection and Preparation	35
3.2.2.2	Data Cleaning	39
3.3	Design Phase	40
3.3.1	System Architecture	41
3.3.2	Flowchart	42
3.3.3	Pseudocode of Selected Algorithm	43
3.3.4	User Interface Design	44
3.3.5	Prototype Implementation	46
3.3.5.1	Software and Hardware Requirements	46
3.4	Evaluation Phase	48
3.5	Gantt Chart	52
3.6	Conclusion	53
CHAPTER FOUR: RESULT AND FINDING		54
4.1	Conceptual Framework	54
4.2	Data Preparation	56
4.3	Classification Model	61