

**Universiti Teknologi MARA**

**Sentiment Analysis on Airline Reviews  
using Naïve Bayes**

**Nurul Sarah Aliessa Che Harun**

**Thesis submitted in fulfilment of the requirements  
for Bachelor of Computer Science (Hons.)**

**Jan 2025**

## **ACKNOWLEDGEMENT**

Alhamdulillah, praises and thanks to Allah for His Almighty and utmost blessings, which enabled the completion of this research within the given timeframe. Above all, special thanks go to my supervisor, Pn. Ummu Fatimah Binti Mohd Bahrin, for her valuable advice and encouragement, enlightening comments that were of utmost importance to the success of this study. She is fully committed to ensuring the accuracy and comprehensiveness of my work. Special thanks also go to my family and friends who always gave me support throughout my family and friends who always gave me support throughout my academic life. This achievement is a testament to Allah's blessings and guidance, and I am deeply grateful for the opportunities that have been afforded to me.

## ABSTRACT

Airlines get a vast number of customer feedback each day but manually analysing them may be time-consuming and incorrect. The unstructured framework of these reviews, along with class inequalities, make it difficult to extract useful information. This study used the Naïve Bayes algorithm to analyse airline reviews and customer feedback, resulting in improved service quality. This may include a preliminary design and implementation evaluation that is further divided into three phases. At the preliminary phase, the literature review was performed, and there was a selection of the dataset composed of 9,210 Airline Reviews from 2020 to 2023 from Kaggle. In the design and implementation phases, TF-IDF was considered in cleaning and preparing the reviews for application, while the construction of a sentiment analysis model was implemented through the use of the Multinomial Naïve Bayes algorithm. A friendly GUI was developed for the ease of reviews analysis. Three splits were taken for testing of the model, out of which 70:30 produced the best results, having AUC (Area Under the Curve) to be 0.89, whereas the accuracy is 84.58%. In order to tackle class imbalance in the dataset, ADASYN oversampling was done and then, afterward, the model tuned to more reliable regarding detecting positive and negative sentiments. The findings proved that this algorithm of Naïve Bayes is appropriate and efficient to analyze reviews provided for airlines. This would really give the opportunity to understand customer feedback clearly, along with faster problems fixation, resulting in an enhanced service quality being developed. This technique in the near future can further be enhanced through complex algorithms by applying it on bigger datasets also, which creates the platform for enhancement in customers' satisfaction and makes airlines smarter too.

# TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL .....	i
STUDENT DECLARATION.....	ii
ACKNOWLEDGEMENT.....	iii
ABSTRACT .....	iv
TABLE OF CONTENTS .....	v
LIST OF FIGURES .....	viii
LIST OF TABLES .....	ix

## CHAPTER 1: INTRODUCTION

1.1 Background Study.....	1
1.2 Problem Statement.....	3
1.3 Objectives.....	4
1.4 Project Scope.....	5
1.5 Project Significance.....	6

## CHAPTER 2: LITERATURE REVIEW

2.1 Sentiment Analysis.....	12
2.1.1 Overview of Sentiment Analysis.....	12
2.1.2 Applications of Sentiment Analysis in Airline Reviews.....	14
2.1.3 Overview of Different Sentiment Analysis Method.....	15
2.3 Naïve Bayes.....	17
2.2.1 Overview of Naïve Bayes.....	17
2.2.2 Mathematical Foundation.....	19
2.2.3 Bernoulli Naïve Bayes.....	21
2.2.4 Multinomial Naïve Bayes.....	21
2.2.5 Gaussian Naïve Bayes .....	22
2.2.6 Advantages of Naïve Bayes Classifier .....	22
2.3 Similar Works of Naïve Bayes Algorithm.....	25
2.4 Airline Reviews .....	30
2.5 Implication of Literature Review.....	31
2.6 Conclusion.....	33

## **CHAPTER 3: RESEARCH METHODOLOGY**

3.1 Overview of Research Methodology .....	34
3.1.1 Detailed of Research Methodology Framework .....	34
3.2 Preliminary Phase .....	39
3.2.1 Literature Review .....	39
3.2.2 Study of Naïve Bayes algorithm .....	40
3.2.3 Airline Review Dataset .....	41
3.3 Design and Implementation Phase .....	41
3.3.1 Data Pre-processing .....	43
3.3.2 System Logical Design .....	46
3.3.3 Flowchart of Sentiment Analysis using Naïve Bayes .....	51
3.3.4 Graphical User Interface (GUI) .....	53
3.3.5 Prototype Implementation .....	54
3.4 Evaluation Phase .....	56
3.5 Conclusion .....	59

## **CHAPTER 4: RESULT AND DISCUSSION**

4.1 System Architecture .....	60
4.2 Program Codes .....	62
4.2.1 Data Pre-processing .....	62
4.2.2 Data Labelling .....	71
4.2.3 Visualizing Sentiments with Word Cloud .....	72
4.2.4 TF-IDF Vectorizer .....	75
4.2.5 Implementation of Naïve Bayes Algorithm .....	76
4.3 User Interface .....	80
4.4 Evaluation Results .....	85
4.4.1 Experiment 1: Data Testing without ADASYN on 70:30 .....	85
4.4.2 Experiment 2: Data Testing without ADASYN on 80:20 .....	89
4.4.3 Experiment 3: Data Testing without ADASYN on 90:10 .....	92
4.4.4 Experiment 4: Data Testing with ADASYN on 70:30 .....	95
4.4.5 Experiment 5: Data Testing with ADASYN on 80:20 .....	98
4.4.6 Experiment 6: Data Testing with ADASYN on 90:10 .....	101
4.4.7 Accuracy Testing .....	104
4.4.8 AUC Score Comparisons .....	105
4.5 Discussion .....	107
4.6 Conclusion .....	108