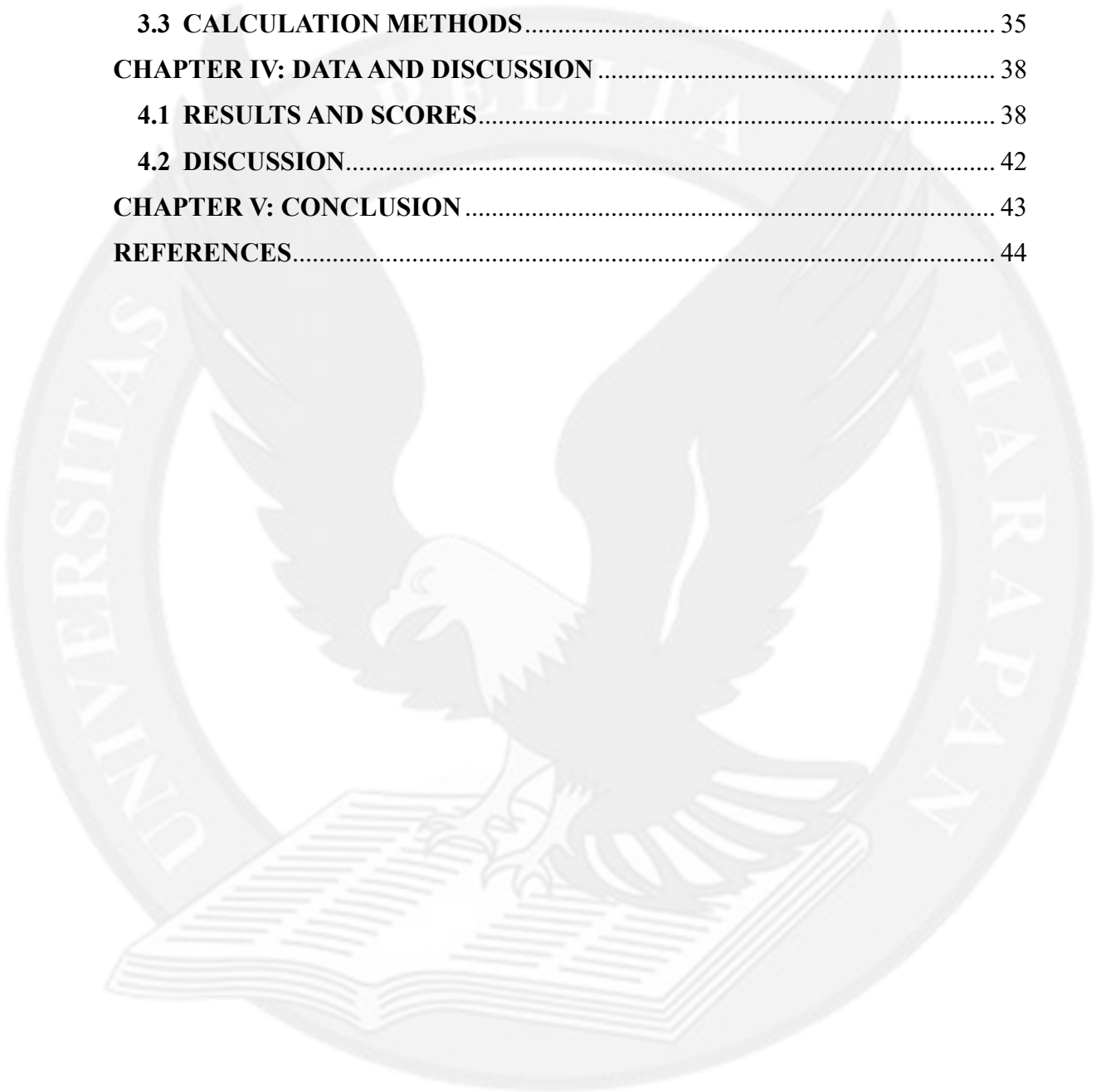


TABLE OF CONTENTS

COVER	
ORIGINALITY STATEMENT OF THE THESIS	
THESIS SUPERVISOR'S APPROVAL	
THESIS EXAMINATION COMMITTEE	
ABSTRACT	v
ABSTRAK	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xi
CHAPTER 1: INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM IDENTIFICATION	3
1.3 PROBLEM LIMITATION	4
1.4 PROBLEM DEFINITION	4
1.5 PURPOSE OF THE STUDY	5
1.6 OUTLINE	5
CHAPTER II: LITERATURE REVIEW	7
2.1 GAUGING AIRLINE SATISFACTION	7
2.2 FACTORS THAT AFFECT AIRLINE SATISFACTION	9
2.3 ENSEMBLE LEARNING TECHNIQUES	14
2.3.1 BAGGING	14
2.3.2 BOOSTING	18
2.3.3 STACKING	21
Chapter III: METHODOLOGY	23
3.1 DATASET	23
3.1.1 DATA PRE-PROCESSING	27

3.1.2 EXPLORATORY DATA ANALYSIS	30
3.2 EVALUATION METHOD	33
3.3 CALCULATION METHODS.....	35
CHAPTER IV: DATA AND DISCUSSION	38
4.1 RESULTS AND SCORES.....	38
4.2 DISCUSSION.....	42
CHAPTER V: CONCLUSION	43
REFERENCES.....	44



LIST OF FIGURES

Figure 1.0.1: the passenger traffic by each region from 2019 to 2024 prediction. Source: (Airports Council International, 2023).	1
Figure 2.0.1: Graph showing the different results in rating of a survey of a restaurant. Source: (Masoud Kamalahmadi, 2023).	8
Figure 2.0.2: A bar graph chart showing the reasons that affects a customer's loyalty towards an airline. Source: (Neringa Slavinskaite, 2023)	9
Figure 2.0.3: bar chart showing the differences in overall rating between genders, with 1 being male and 2 being female. Source: (Ashok K. Singh, 2019).	10
Figure 2.0.4: bar chart showing the age differences in rating the airline satisfaction, with 1 being under 18, 2 being 18-24, 3 being 25-34, 4 being 35-44, 5 being 45-54, 6 being 55-64, 7 being 65 and over, and 9 being multiple responses. Source: (Ashok K. Singh, 2019).	11
Figure 2.0.5: two separate charts showing delay time and the emotion level of the passengers before and during delay. Source: (Quan Shao, 2021).	11
Figure 2.0.6: the process of how bagging works. The “Bootstrap” is where subsets of a dataset which are randomly selected with replacement, and then put into a classifier or regression to each subset. Source: (Awan, 2023).	14
Figure 2.0.7: diagram demonstrating how Random Forest works. Source: (Chaudhary, 2023).	15
Figure 2.0.8: demonstration of how Decision Tree method works in understanding risks to prevent heart attack. Source: (Navlani, 2023).	16
Figure 2.0.9: a flowchart showing the process of Boosting method. Firstly, the training dataset is split into three equal-sized subsets with replacement, which are then used for each models independently and in parallel, before finally combining each results through averaging or voting to get the final result. Source: (Budu, 2023).	18
Figure 2.0.10: a flowchart showing the process of Stacking in ensemble learning. Source: (Bantu, 2020).	21
Figure 3.0.1: dataset info for the "training" dataset for this study.	22

Figure 3.0.2: dataset attributes of the dataset "test".	23
Figure 3.0.3: the data attributes of the dataset "training" after the fix. The “unnamed column” together with “id” column as seen on figure 3.0.1 has been removed, and most datatypes from number 6 to 19 has been changed into “category” data type to better fit for inputting on ensemble learning.	26
Figure 3.0.4: dataset attributes of the dataset "test" after the fix. The “unnamed column” together with “id” column as seen on figure 3.0.2 has been removed, and most datatypes from number 6 to 19 has been changed into “category” data type to better fit for inputting on ensemble learning.	26
Figure 3.0.5: the list of total missing values on each column of "training" dataset. Note the number of missing values in Arrival Delay in Minutes column.	27
Figure 3.0.6: the list of total missing values on each column of "test" dataset. Note the number of missing values in the Arrival Delay in Minutes column.	27
Figure 3.0.7: the list of total missing values on the "training" dataset after the fix.	28
Figure 3.0.8: the list of total missing values on the "test" dataset after the fix.	28
Figure 3.0.9: Pie chart showing the overall number of neutral or dissatisfied, and satisfied passengers within the dataset, overall “training” dataset is almost near balanced among both level of satisfaction.	29
Figure 3.0.10: correlation heatmap showing the relationship of each column with one another in ”training” dataset. Note the strong correlation between the columns departure delay and arrival delay.	29
Figure 3.0.11: the correlation heatmap showing the relations between quantitative values within the dataset. Note the strong correlation between departure delay in minutes with the arrival delay in minutes.	30
Figure 3.0.12: bar chart showing the differences of satisfaction levels between the two genders.	30
Figure 3.0.13: bar chart showing the total number of satisfied and neutral or dissatisfied passengers, split between loyal and disloyal type of customer.	31
Figure 3.0.14: box chart and histogram plot showing therelations of flight distance and inflight entertainment with satisfaction levels.	31
Figure 3.0.15: confusion matrix and what each value signifies.	32

Figure 3.0.16: an ROC curve, where scores above 0.5 and higher is accurate, while scores under 0.5 is less accurate. Values larger than 0.5 also indicates that model has an ability to discriminate.	33
Figure 3.0.17: a bar chart of Feature Importance. The higher the score is, the more it will affect the overall model scoring.	33
Figure 4.0.1: confusion matrix and precision, recall, f1-score, and support scores of Decision Tree ensemble learning method.	36
Figure 4.0.2: Feature importance for Decision Tree ensemble learning method.	36
Figure 4.0.3: ROC curve for Decision Tree ensemble learning method.	37
Figure 4.0.3: confusion matrix and precision, recall, f1-score, and support scores for the Random Forest ensemble learning method.	37
Figure 4.0.4: features importance of Random Forest ensemble learning method.	37
Figure 4.0.6: ROC curve of Random Forest ensemble learning method.	38
Figure 4.0.7: the confusion matrix and precision, recall, f1-score, and support scores for the Boosting ensemble learning method.	38
Figure 4.0.8: the features importance chart of the Boosting ensemble learning method.	39
Figure 4.0.9: ROC Curve of the Boosting ensemble learning method.	39
Figure 4.0.10: confusion matrix and precision, recall, f1-score, and support scores for Stacking ensemble learning method.	39
Figure 4.0.11: the features importance chart of the Stacking ensemble learning method.	39
Figure 4.0.12: ROC curve of Stacking ensemble learning method.	40