# CHAPTER 1

# INTRODUCTION

## 1.1    Background

As of 2023, total passengers boarding airlines has increased by 30.1% compared to last year, showing strong recovery from COVID-19 pandemic and will continue to see a strong growth trend in the future (Airlines IATA, 2023), with the chart below from another source predicting that global passenger traffic will fully recover by 2024 and may reach 9.4 billion passengers Figure 1.1.
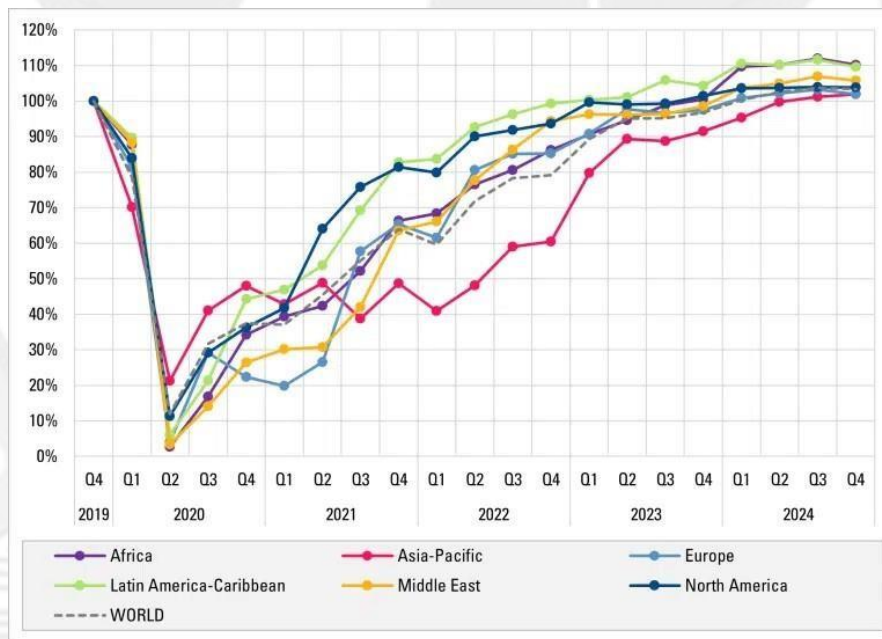


Figure 1.0.1: the passenger traffic by each region from 2019 to 2024 prediction. Source: *(Airports Council International, 2023).*

With the resurgence of airline traffic post-pandemic, airline industry will need to recover the losses sustained during the lockdown era and the competition to obtain as many passengers as possible will be challenging as industries are struggling to survive

and recover (Jaap Bouwer, 2021). To stay afloat and competitive, airliners must attract potential customers to them while building customer loyalty and recommendation, and one of the best ways to do so would be to increase customer satisfaction (Yanmin Dong, 2021). With the complexity of identifying and analyzing the overall customer satisfaction, airline industries have turned to machine learning, specifically ensemble learning for making complex calculations and reporting on customer satisfaction analysis.

Machine learning in its basic definition, describes the ability of a system to learn from given data related to analytics and solving given problems, which works by slowly learning meaningful patterns and relationships between pieces of data through examples and observations (Christian Janiesch, 2021). Combining multiple machine learning algorithms will combine the output methods to perform more complex calculations for better results which is called Ensemble Learning (Zhou, 2009).

Past studies on analyses of a multi-dimensional problem have seen higher prediction accuracy using ensemble learning than single-based machine learning techniques (Theddeus T. Akano, 2022), with many literature reviews on various ensemble learning techniques (Xibin Dong, 2020). However, despite the strengths, ensemble learning process still have its weaknesses to be aware of, which despite several strategies and techniques still have limitations in terms of generalization, training difficulties, and more (Erdal Tasci, 2021).

Thus, this thesis which is titled "Comparison of Several Ensemble Models for

Airline Customer Satisfaction", will firstly explore what factors will increase airline satisfaction from past studies, and then applying them into the ensemble learning study and comparison. The study will then take an airline dataset which will then be preprocessed and cleaned before using it for both model training and testing, and then building each ensemble learning methods for testing to obtain performance results, which consists of ensemble learning performance metrics, an ROC and performance curve, and a confusion matrix, with which the results are then compared side-by-side with each ensemble learning methods to determine which among them have the best performance.

## 1.2    Problem Identification

From the background provided, the main issues that will be discussed on this paper will be thus:

1. What are the current known ensemble learning methods to be used for comparative analysis in this study?

2. What are the parameters and values from the chosen dataset that will be used for the comparative analysis between all chosen methods?

3. What will be the expected results of each ensemble learning method for comparative analysis?

4. How will the ensemble learning methods be built and tested, and how will the constructed model be used in this sense?

5. How will the airline satisfaction survey responses be quantified and converted into usable data for the ensemble learning methods?

## 1.3    Problem Limitation

With the problems for this study identified, the next step will be to determine what focus will this study be, thus the limitations in this thesis are thus:

1. The number of ensemble learning techniques to be used for this study.

2. The factors of airline satisfaction will be the values and parameters in a dataset which will be used for the ensemble learning study.

3. The expected results will be the accuracy and performance scores in the form of numeric values, which are supplied with charts supporting it.

4. The ensemble learning methods will be built and tested on Jupyter Notebook, using Python programming language, with NumPy, Seaborn, and Pandas plugin to facilitate data gathering, and model building.

5. The survey responses which will be used for the data collection will have data values in either Boolean or numeric values with no open-ended questions.

## 1.4    Problem Definition

With the problems listed above, this study will face several challenges:

1. Which methods of ensemble learning will be chosen for this study, and how to implement it for predicting airline satisfaction.

2. Determining what factors of airline satisfaction is needed as the values and parameters for the ensemble learning for this study.

3. How to quantify and calculate the performance of the airline satisfaction prediction results, and how to present it.

4. Designing ensemble learning method tests that will accept data from airline satisfaction survey results, and getting accurate predictions.

5. How will the airline satisfaction ratings survey be designed in such a way that quantitative data can be obtained for use in ensemble learning methods.

## 1.5    Purpose of The Study

While the main goal of this study is to provide a comparison between the chosen ensemble learning methods on the overall performance and highlighting the score differences among them, this study shall also act as a guide for both new and experienced machine learning practitioners on the problem of prediction in airline satisfaction issue, such as outlining what factors of airline satisfaction is to be taken into account for data input for the ensemble learning machine, and the methodology in building an ensemble learning study for future studies.

## 1.6    Outline

To outline the flow of this thesis for better focus on certain parts of the study objectives, the outline of this paper will be the organized as the following below:

## CHAPTER I. INTRODUCTION

This section will focus on the background of this study to on airline industry and the basics of machine learning and ensemble learning supported by references, with

the problem identification, formulation, and understanding the limits, together with study objectives, and outline of the whole study to show the structure of this thesis.

**CHAPTER II.  STATE OF THE ART**

This section will investigate the factors that affect the customer satisfaction in an airline, guideline in gauging customer satisfaction, and then learning more about the ensemble learning itself, its techniques, and methodologies, which will all be supported by references provided at the appendix.

**CHAPTER 3. METHODOLOGI**

This section will discuss the overall methodology to be used in this study, such as the data gathering, cleaning, and normalization, brief explanations on how the ensemble learning program will work, and the calculation methods to be used on this study, and the workflow of the study itself.

**CHAPTER 4. DATA AND DISCUSSION**

This section will show the results of the ensemble learning study done with the methodology outlined from the previous chapters, complete with images to show the overall passenger satisfaction, and the accuracy scores of each method, and a subsection for discussing and analyzing the outputs of the study.

**CHAPTER 5.  CONCLUSION**

This final section will summarize the overall study results with conclusions and possible recommendations to improve future studies and results.