# Abstract

Flight price prediction is a challenging task that involves forecasting airline ticket prices based on various factors such as historical data, seasonality, demand, and other external influences. Accurate price prediction can be highly beneficial for travellers, airlines, and travel agencies in making informed decisions related to ticket purchasing, revenue management, and pricing strategies.

This project aims to develop a flight price prediction model using machine learning techniques. The dataset used for training and testing the model consists of historical flight data, including features such as departure and arrival locations, date and time of travel, airline carrier, and other relevant information. The dataset is pre-processed to handle missing values, outliers, and undergoes feature engineering to extract meaningful insights.

Multiple machine learning algorithms are implemented and evaluated to determine the most suitable approach for flight price prediction. These include regression models, time series analysis, and ensemble methods. The models are trained on a subset of the dataset and evaluated using appropriate metrics such as accuracy score, mean absolute error (MAE), or root mean square error (RMSE), depending on the algorithm.

In addition to traditional machine learning techniques, advanced methods like deep learning and neural networks may also be explored to capture complex patterns and dependencies within the data. Techniques such as transfer learning and feature selection are employed to enhance model performance and reduce computational complexity.

The trained flight price prediction model can be integrated into a user-friendly application or website where travellers can input their desired travel details and receive accurate price estimates. The model's performance can be continuously monitored and improved over time with the incorporation of additional data.

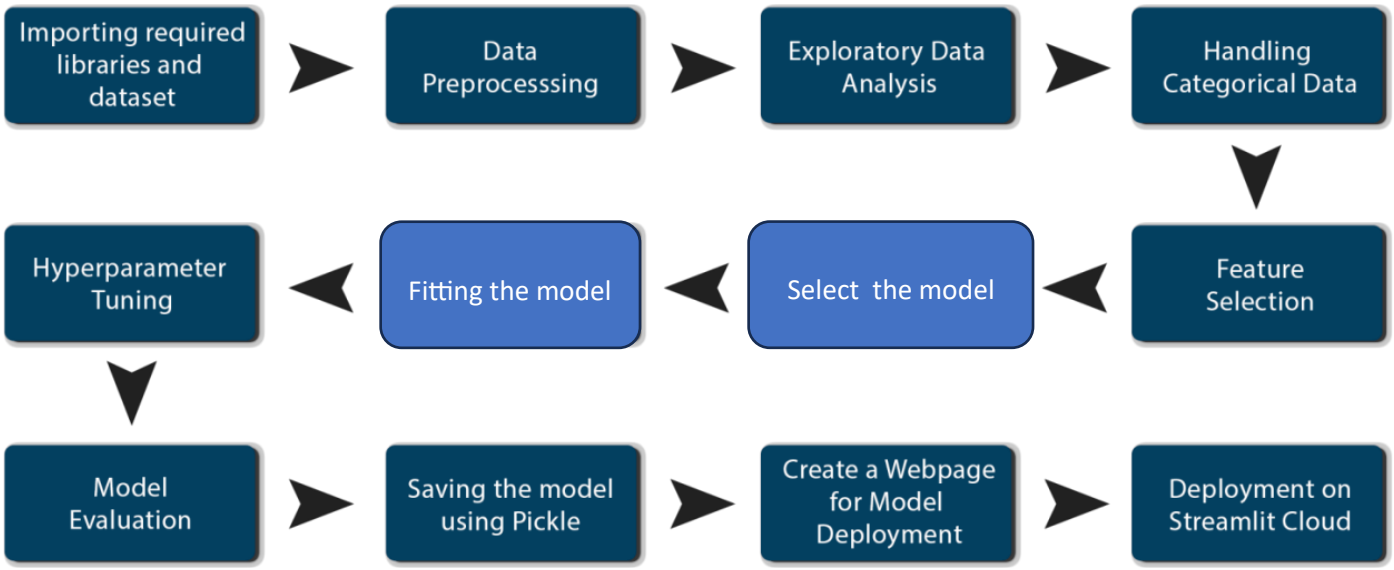
Overall, this project aims to contribute to the travel and tourism industry by developing a reliable and accurate flight price prediction model that supports both travellers and industry professionals in making data-driven decisions related to flight bookings and pricing strategies.

# 1. Introduction

### 1.1 **Why This Architecture Design Document?**

The main objective of the Architecture Design Document is to provide a clear understanding of the internal logic and structure of the **Flight Price Prediction** system. This document is designed in a way that enables developers to directly implement code after reading the description of each module. It serves as a blueprint that bridges the gap between system design and development, ensuring consistency, clarity, and efficiency throughout the implementation process

# 2. Architecture



# 3. Architecture Design

3.1 Data Collection

The data for this project was collected from **Kaggle**, a popular platform for data science datasets.

### 3.2 Data Description

The flight price dataset used in this project contains historical data related to airline flights and their corresponding prices. It includes various features that are important for predicting flight ticket prices. The dataset consists of **10,683 rows** and **11 columns**. The key columns in the dataset include:

- **Airline** – Name of the airline carrier
- **Date_of_Journey** – The date of the flight
- **Source** – Departure city
- **Destination** – Arrival city
- **Stops** – Number of stops between source and destination
- **Dep_Time** – Scheduled departure time
- **Arrival_Time** – Scheduled arrival time
- **Price** – Final ticket price (target variable)

### 3.3 Data Preprocessing

The following steps were performed to prepare the data for training:

- Checked the **data types** of all features using df.info()
- Identified and **removed null values**, as they can negatively affect model accuracy
- Performed **Label Encoding** on categorical features to convert them into numeric form
- Analysed the **distribution** of features to interpret their importance in prediction

At this stage, the data is clean and ready to be used for training a machine learning model.

### 3.4 Modelling Process

After preprocessing, the dataset was visualized to gain insights, and then **split into training and testing sets**. Various machine learning algorithms were applied to the training data to identify the most effective model for predicting flight prices.