FLIGHT DELAY PREDICTION FOR AVIATION INDUSTRY USING MACHINE LEARNING.

# INTRODUCTION

A flight delay is said to occur when an airline lands or takes off later than its scheduled arrival or departure time respectively. Conventionally if a flight's departure time or arrival time is greater than 15 minutes than its scheduled departure and arrival times respectively, then it is considered that there is a departure or arrival delay with respect to corresponding airports. Notable reasons for commercially scheduled flights to delay are adverse weather conditions, air traffic congestion, late reaching aircraft to be used for the flight from previous flight, maintenance and security issues. One of the key business issues that airlines face is that the vital prices that are related to flights being delayed because of natural occurrences and operational shortcomings that is an upscale affair for the airlines, making issues in scheduling and operations for the end users therefore inflicting unhealthy name and client discontent. As we all know that we have a tendency to not get the flight delay before departure as customers of the Airline Company neither the airline company ground staff gets the airline delay prediction supported varied conditions. We can use it to predict if a flight carrier will have a departure delay and hence try to avoid that from happening.

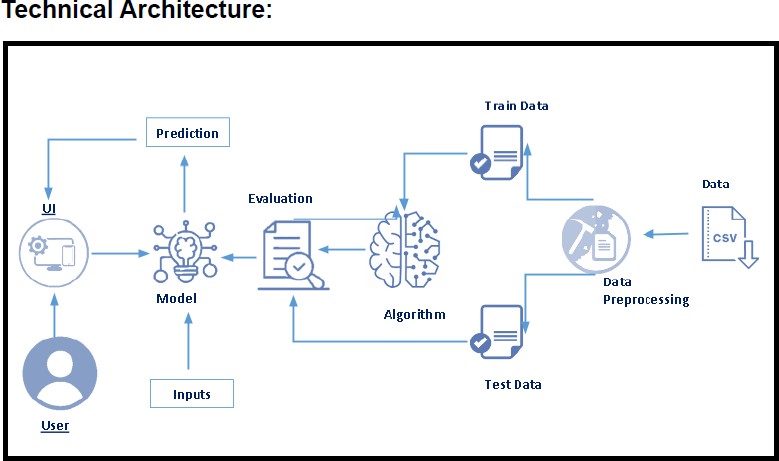
This will prevent the customers as well as the airlines to avoid any losses, whether in their time or business.

* 1. Overview

Analysis of flight delay, therefore, has become a popular research area. Various researchers used different techniques of machine learning and data mining to conduct the investigation. They were interested in different aspects such as airport facility location, weather condition, and airport capacity.

* 1. Purpose

Therefore, predicting flight delays can improve airline operations and passenger satisfaction, which will result in a positive impact on the economy. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.



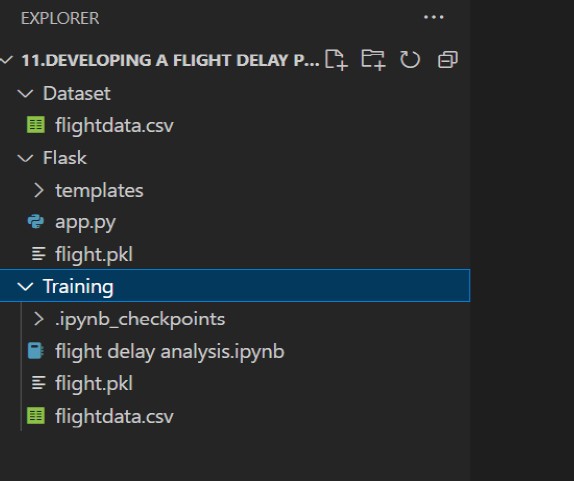
# Project Flow:

* + - User interacts with the UI to enter the input.
    - Entered input is analysed by the model which is integrated.
    - Once model analyses the input the prediction is showcased on the UI To accomplish this, we have to complete all the activities listed below,
    - Define Problem / Problem Understanding
      * Specify the business problem
* Business requirements ○

Literature Survey

* Social or Business Impact.
  + - Data Collection & Preparation
      * Collect the dataset
* Data Preparation
  + - Exploratory Data Analysis
      * Descriptive statistical
* Visual Analysis
  + - Model Building
      * Training the model in multiple algorithms
* Testing the model
  + - Performance Testing & Hyperparameter Tuning
      * Testing model with multiple evaluation metrics
* Comparing model accuracy before & after applying hyperparameter tuning
  + - Model Deployment
      * Save the best model
* Integrate with Web Framework
  + - Project Demonstration & Documentation
      * Record explanation Video for project end to end solution
* Project Documentation-Step by step project development procedure

# Project Structure



* + - We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.
    - flight.pkl is our saved model. Further we will use this model for flask integration.
    - Training folder contains a model training file

Problem Definition & Design Thinking

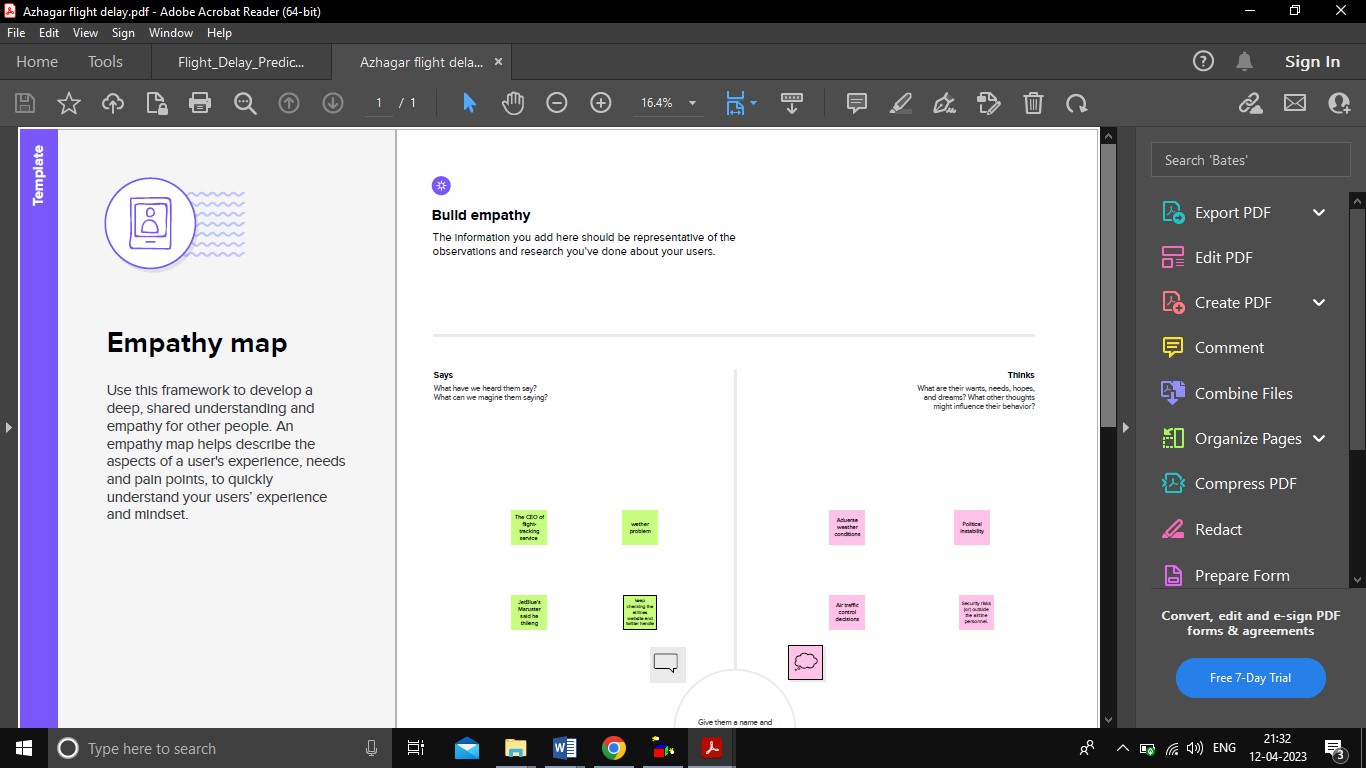
OVER the last twenty years, air travel has been increasingly preferred among travelers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. An increase in air traffic growth has also resulted in massive levels of aircraft delays on the

ground and in the air. These delays are responsible for large economic and environmental losses. According to, taxi-out operations are responsible for 4,000 tons of hydrocarbons, 8,000 tons of nitrogen oxides and 45,000 tons of carbon monoxide emissions in the United States in 2007. Moreover, the economic impact of flight delays for domestic flights in the US is estimated to be more than $19 Billion per year to the airlines and over $41 Billion per year to the national economy In response to growing concerns of fuel emissions and their negative impact on health, there is active research in the aviation industry for finding techniques to predict flight delays accurately in order to optimize flight operations and minimize delays.

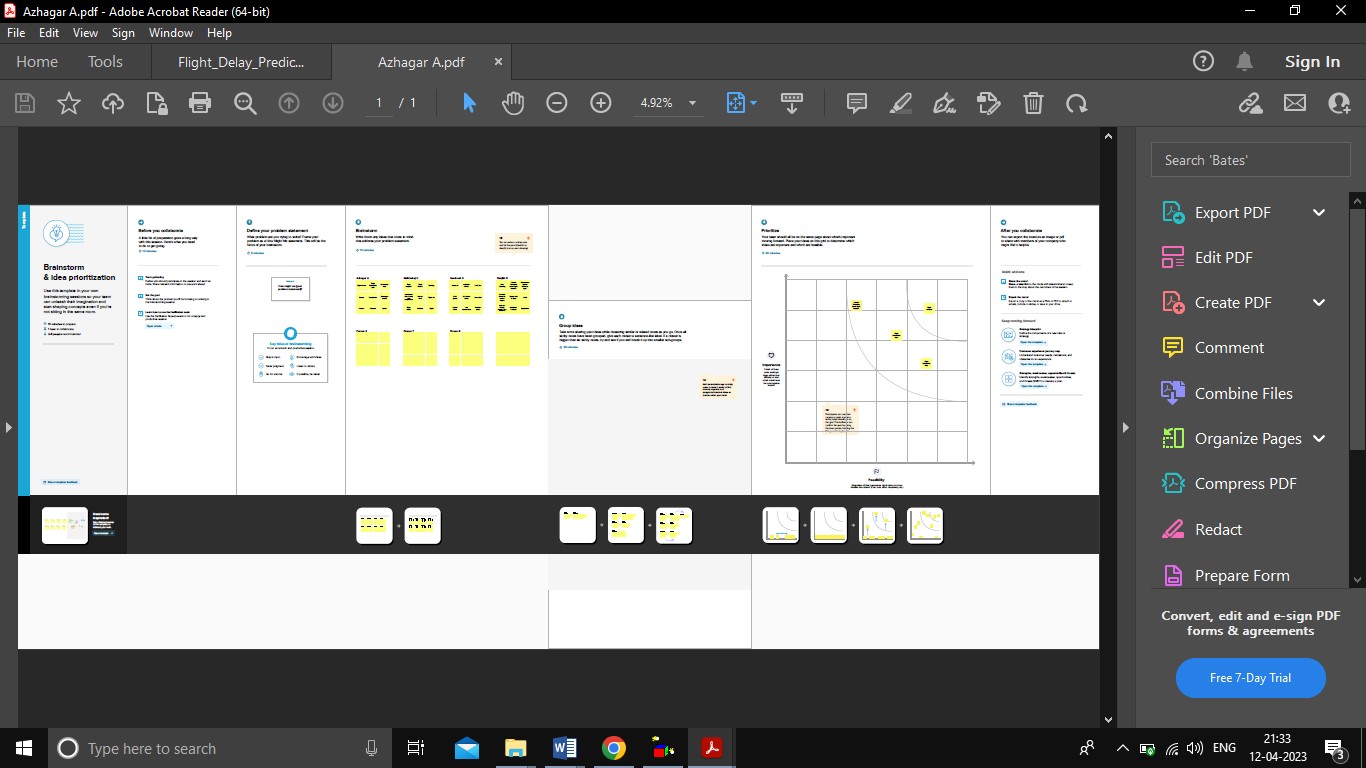
Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifier with logistic regression and a simple neural network for various figures of merit.

Finally, it will be integrated to web based application

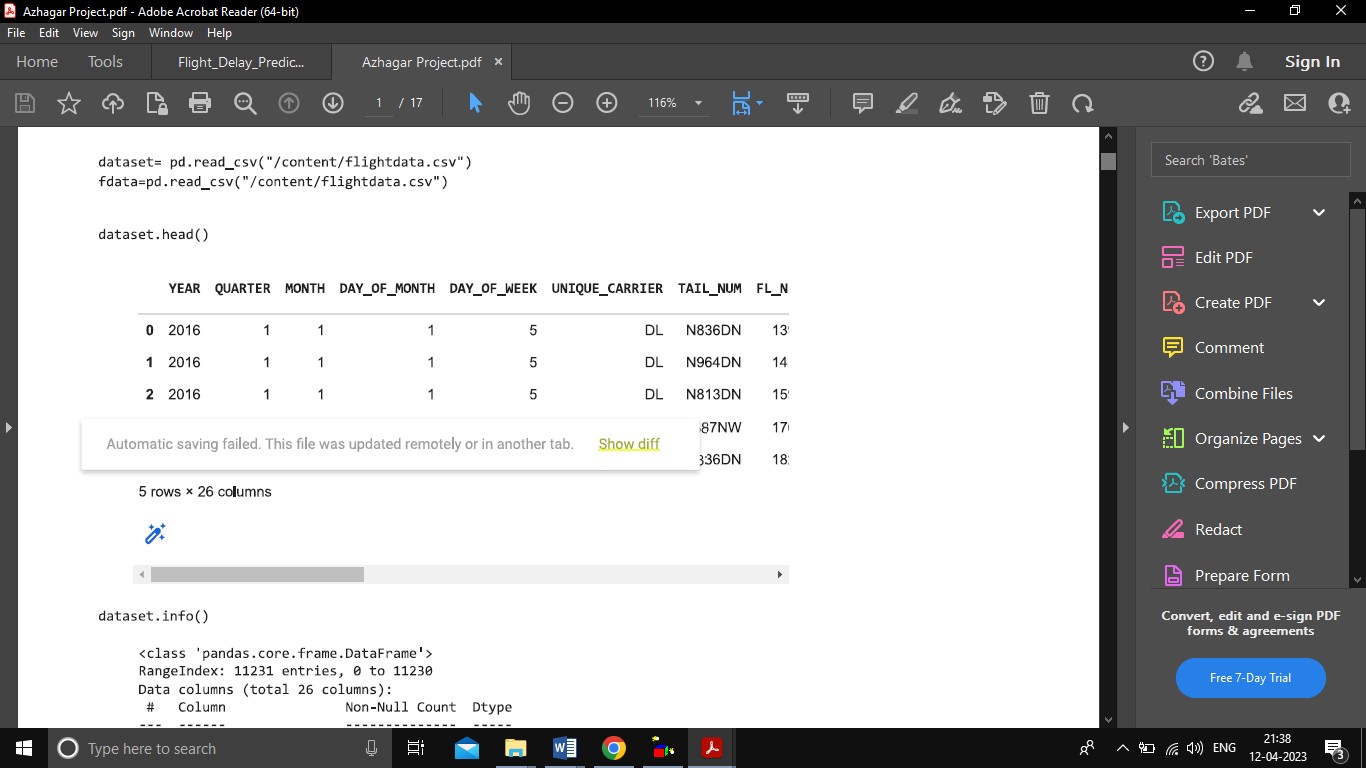
* 1. EMPATHY MAP

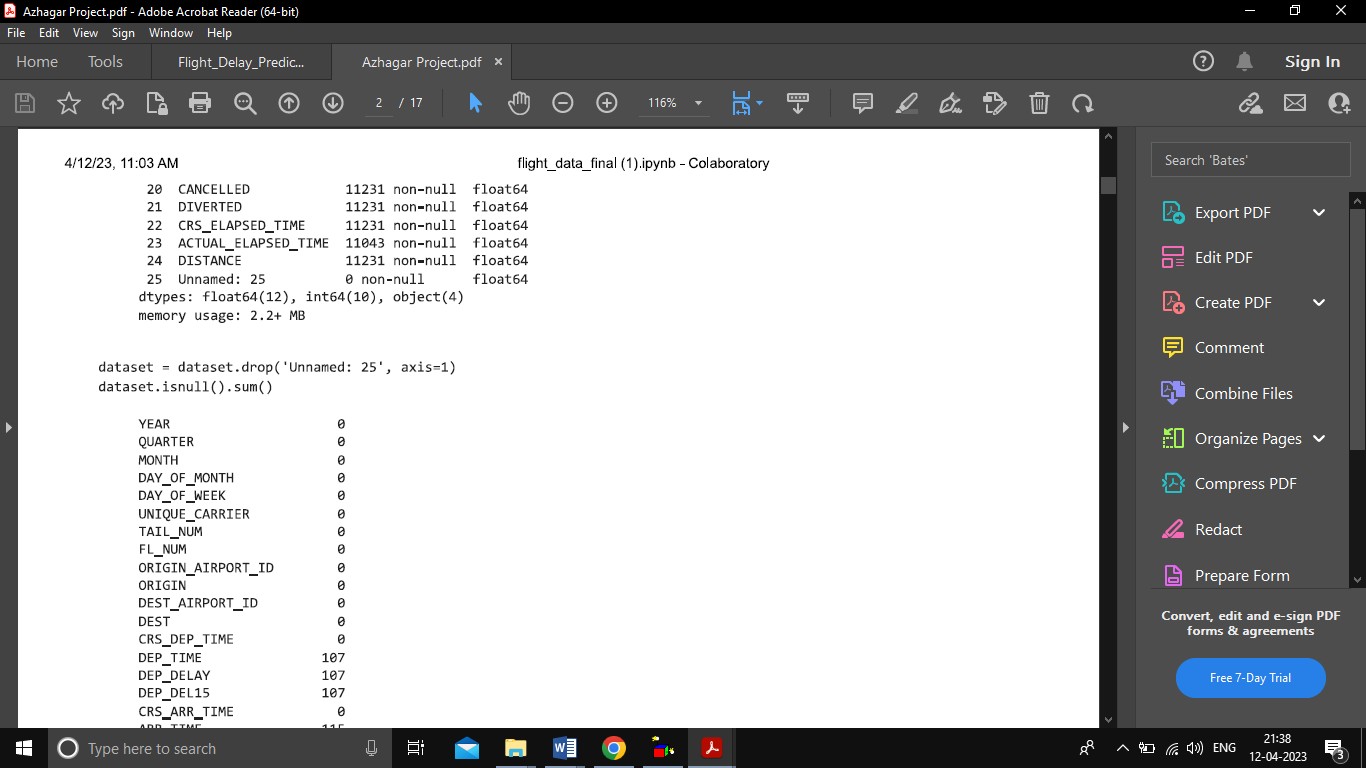


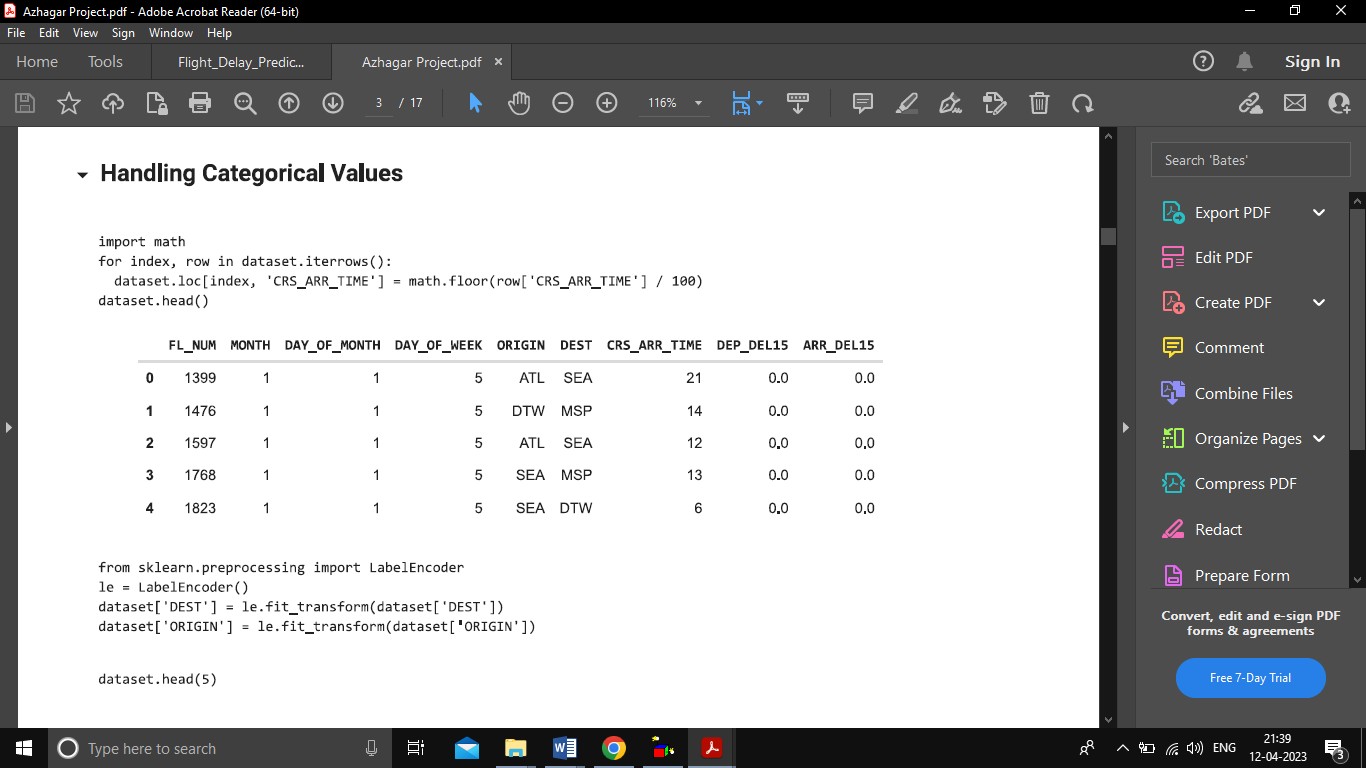
* 1. IDEATION AND BRAINSTORMING MAP

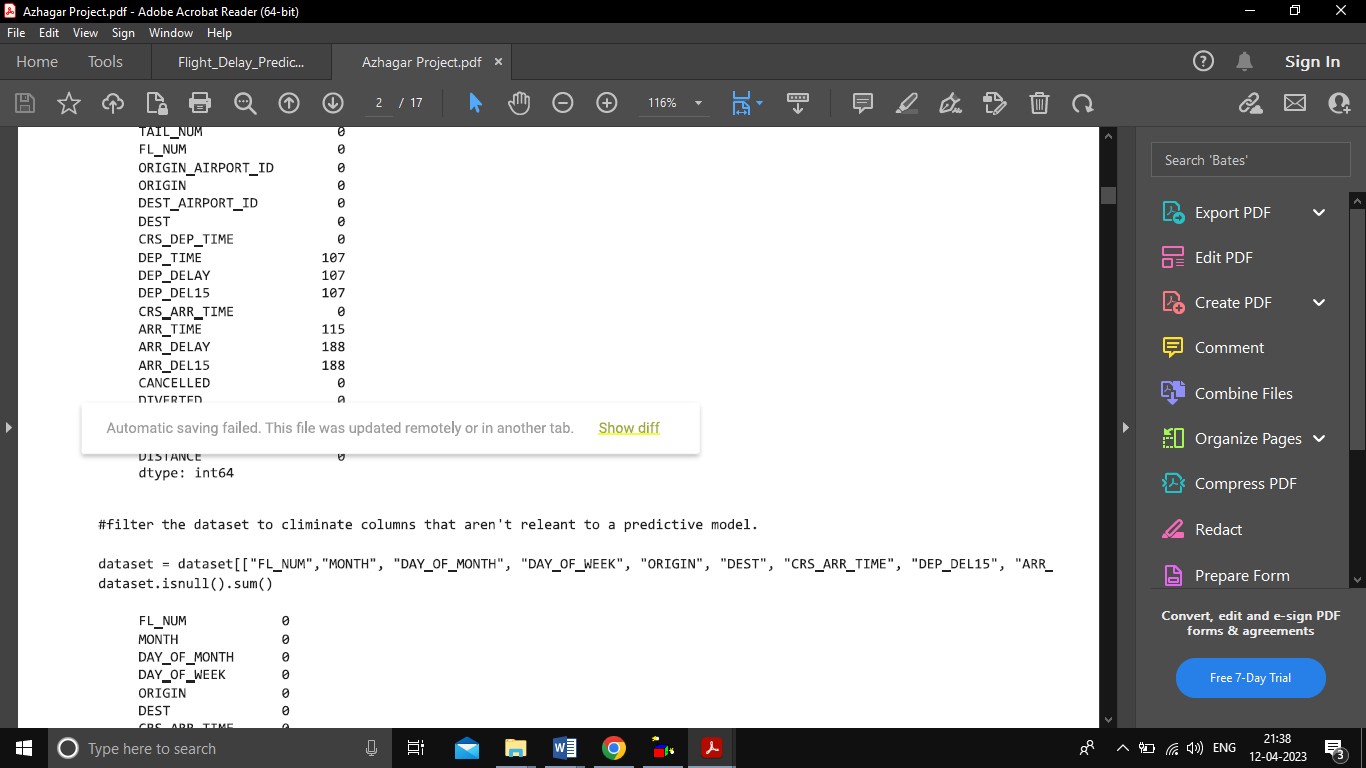


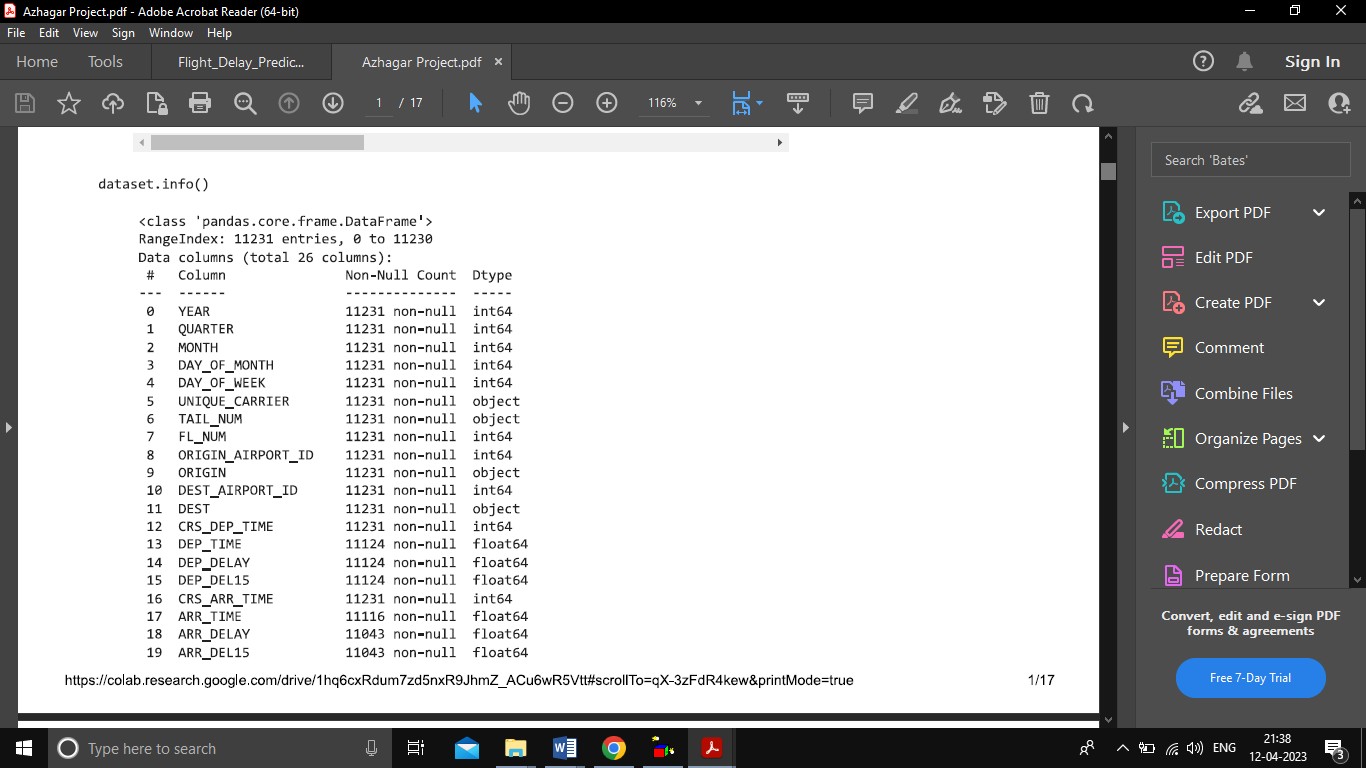
1. RESULT

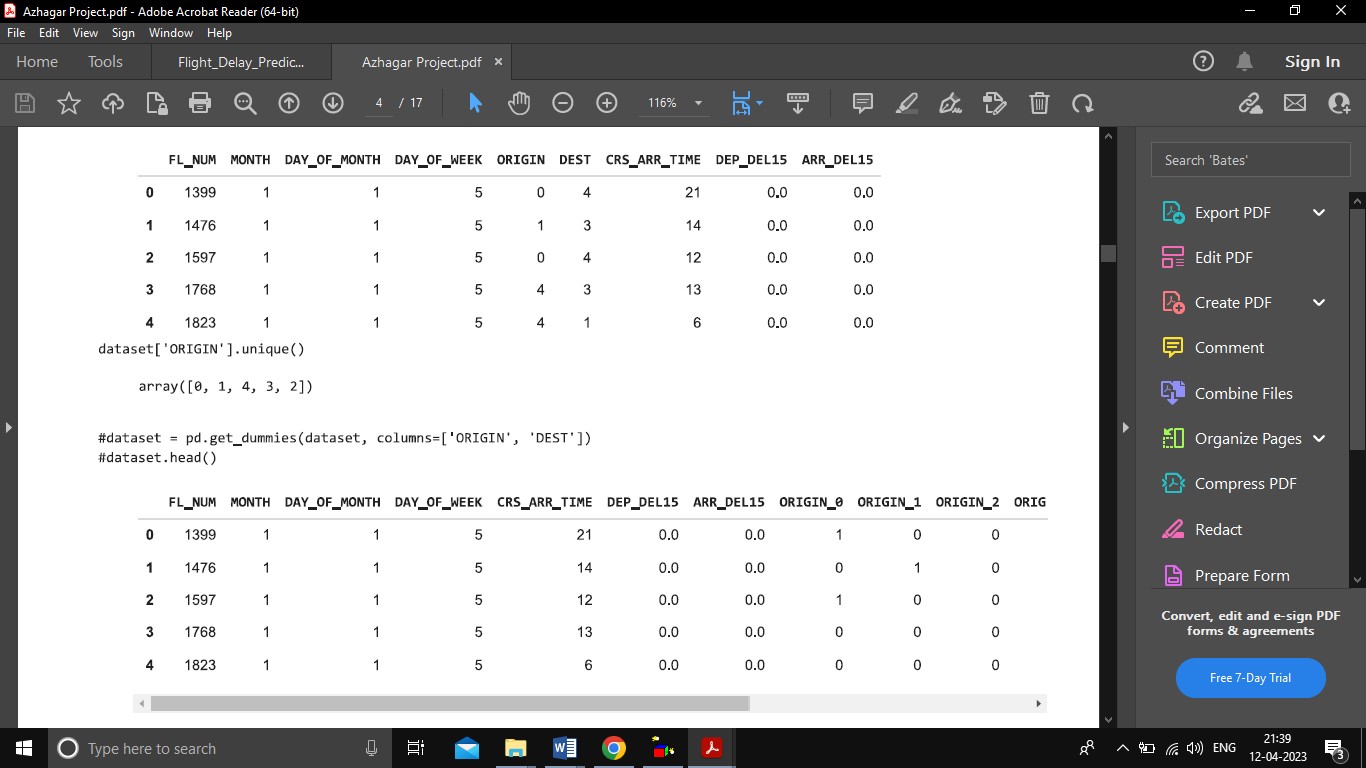


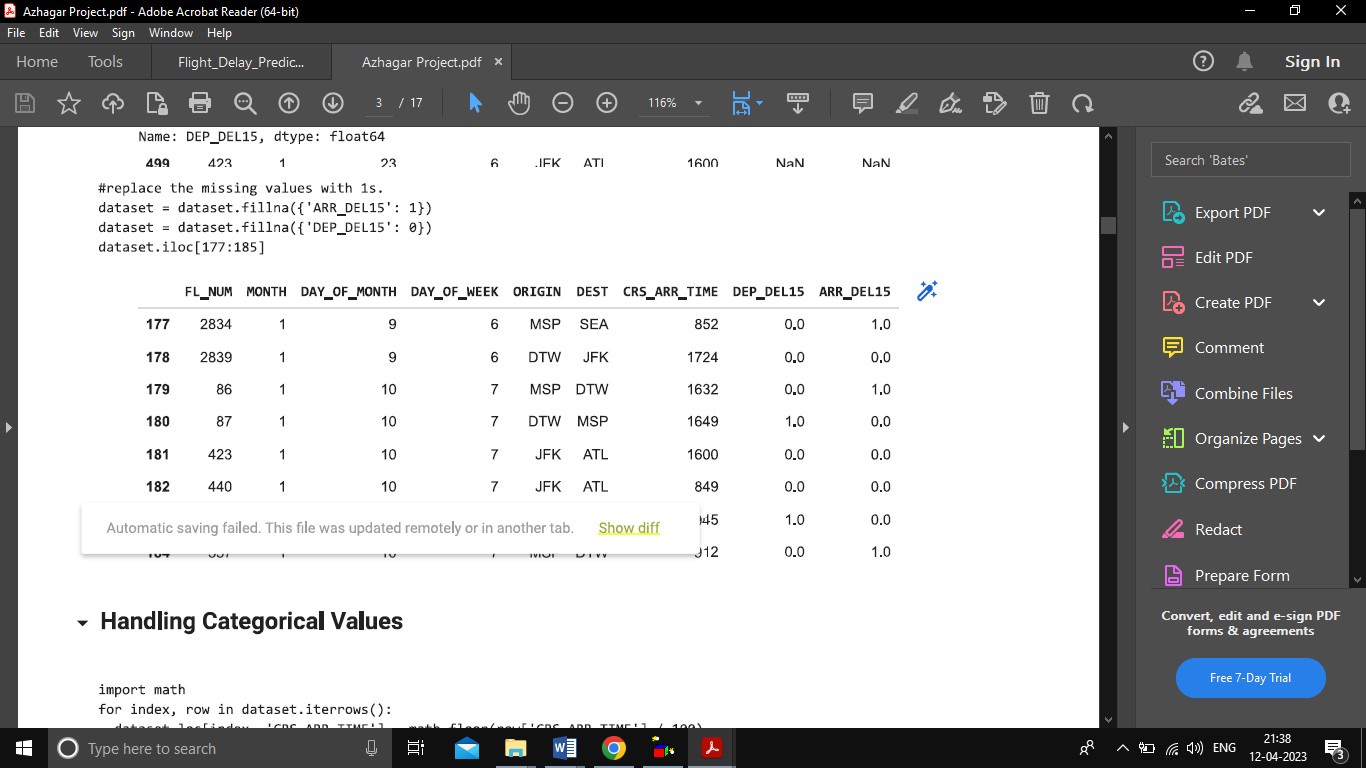


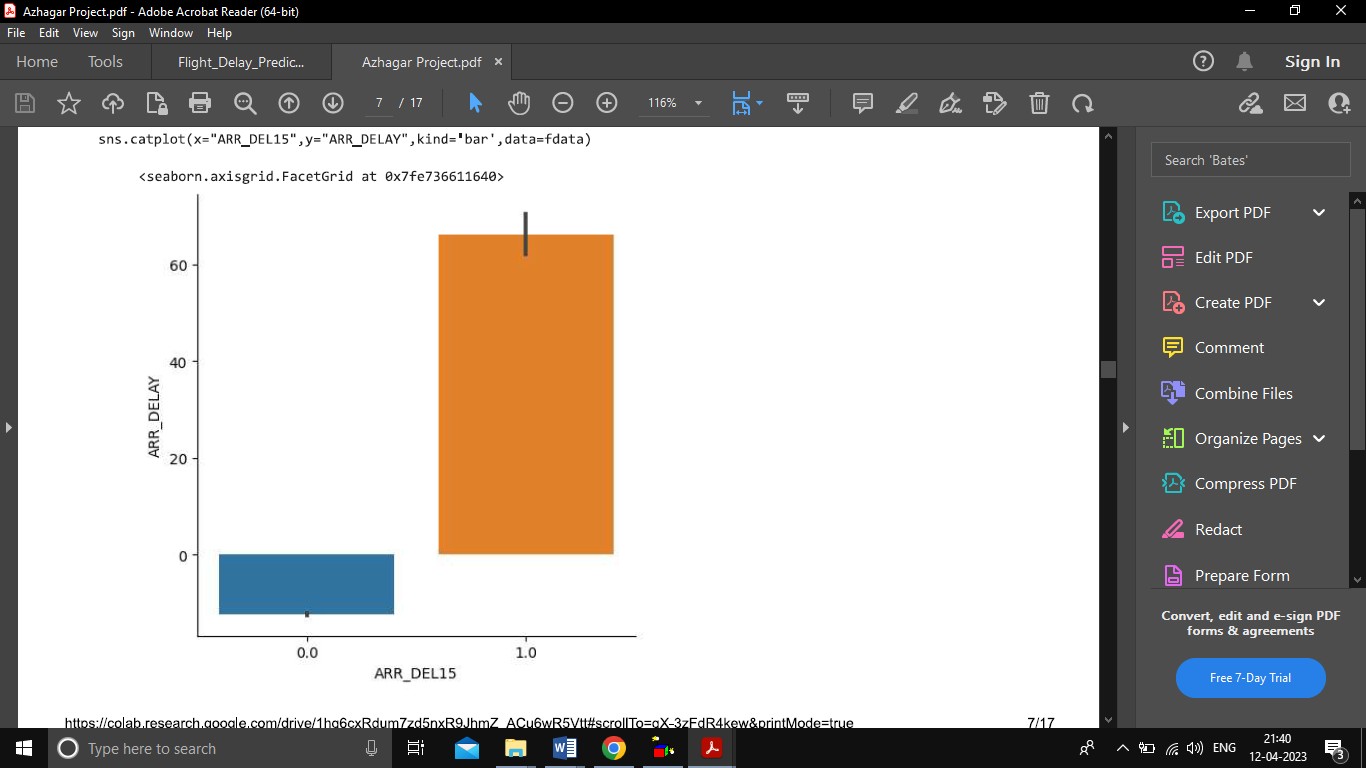


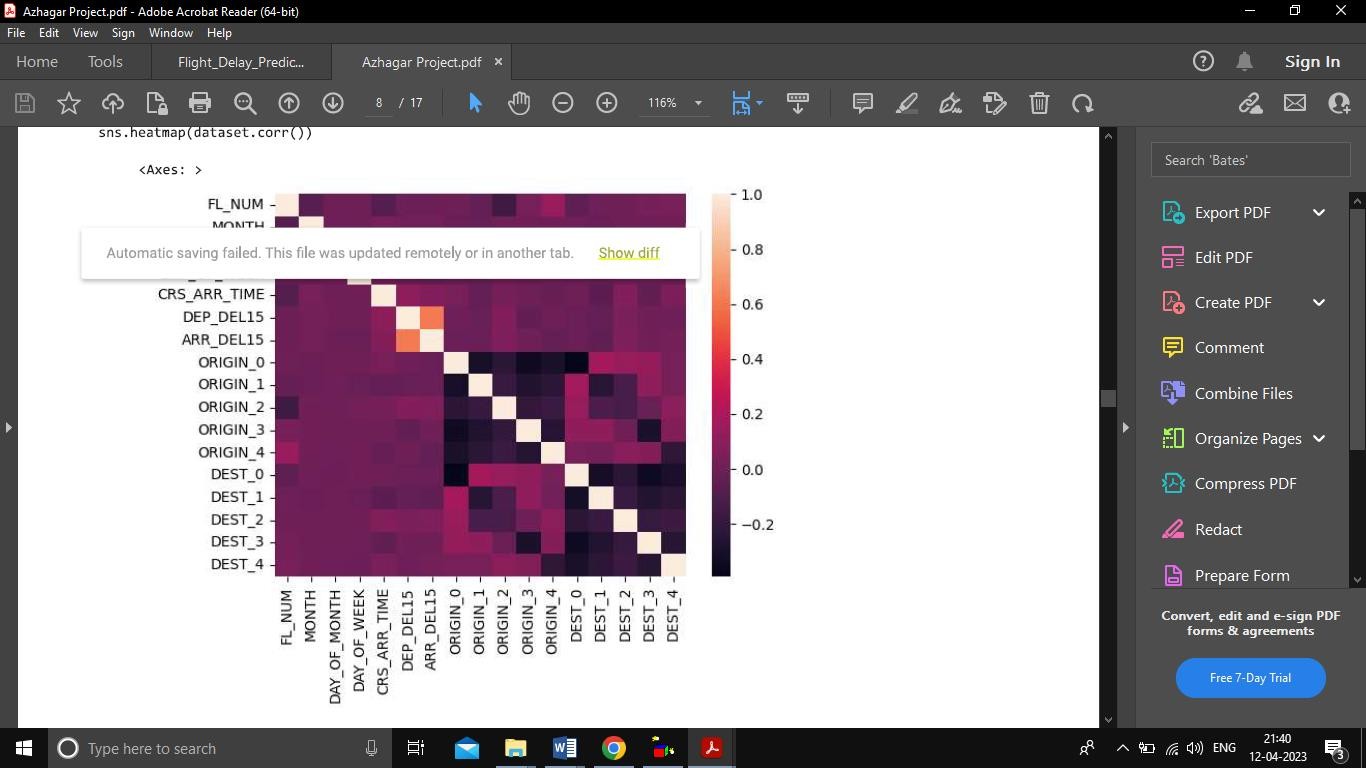


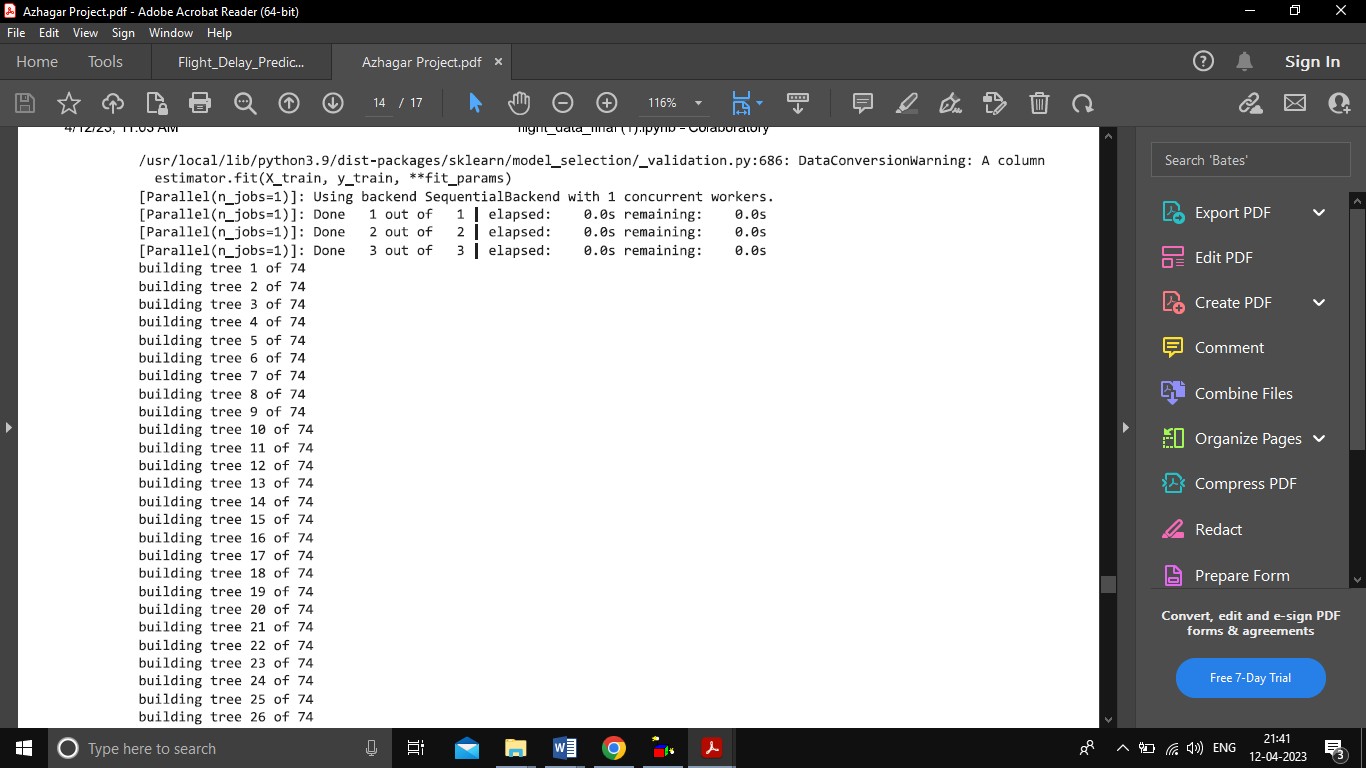


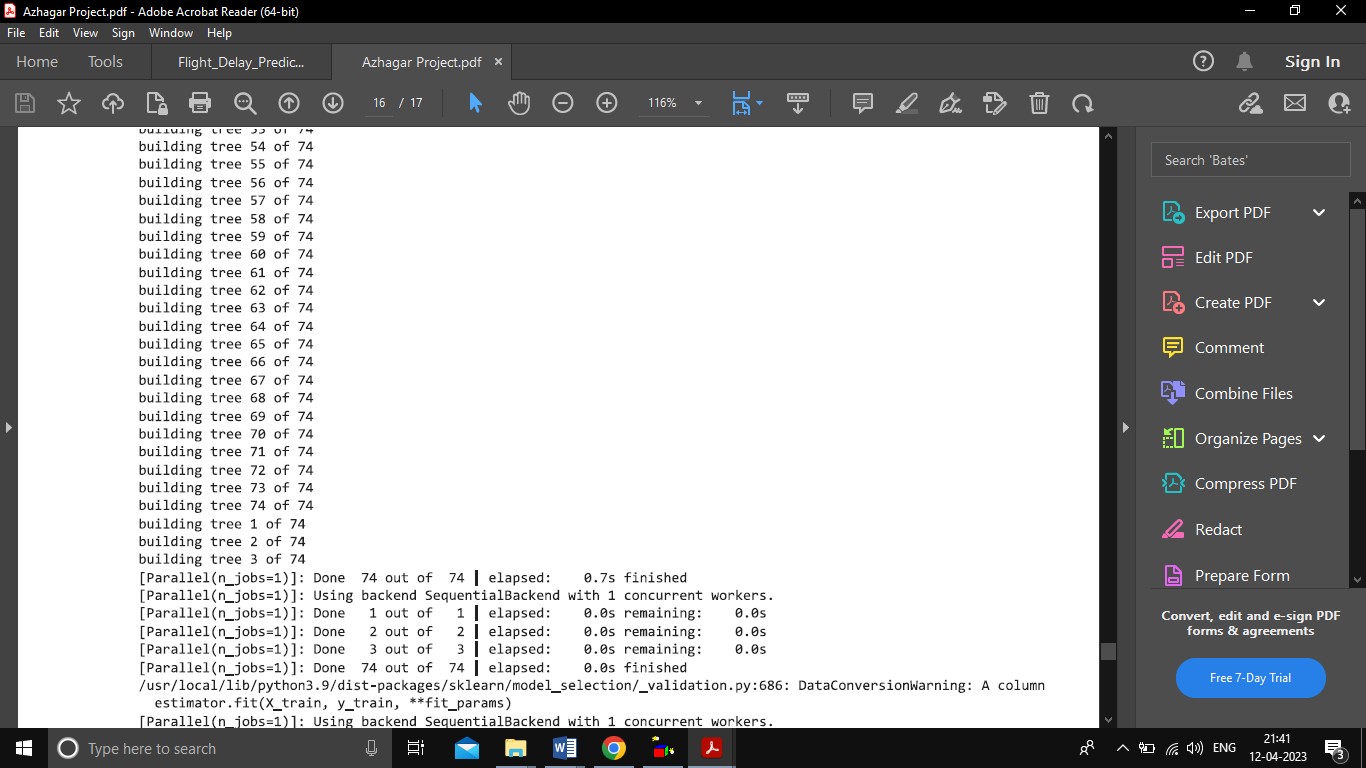


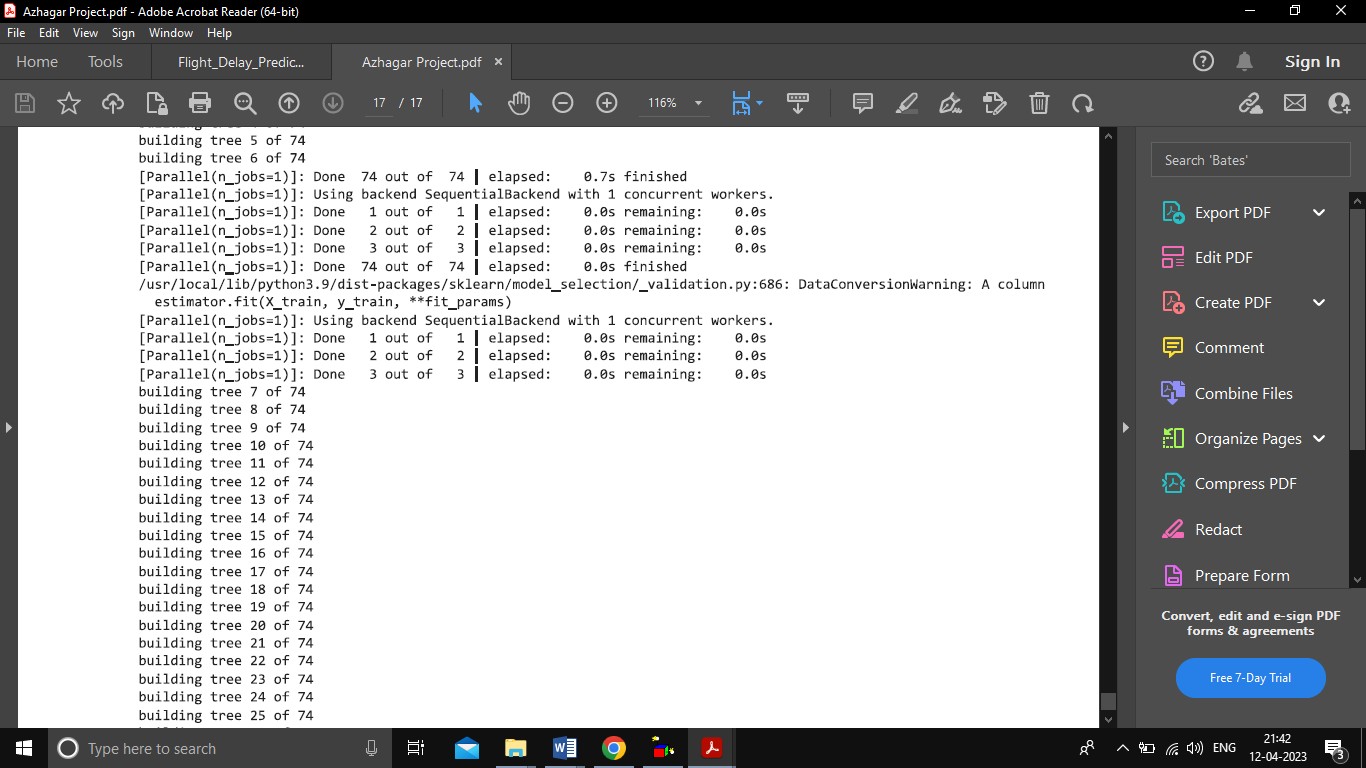


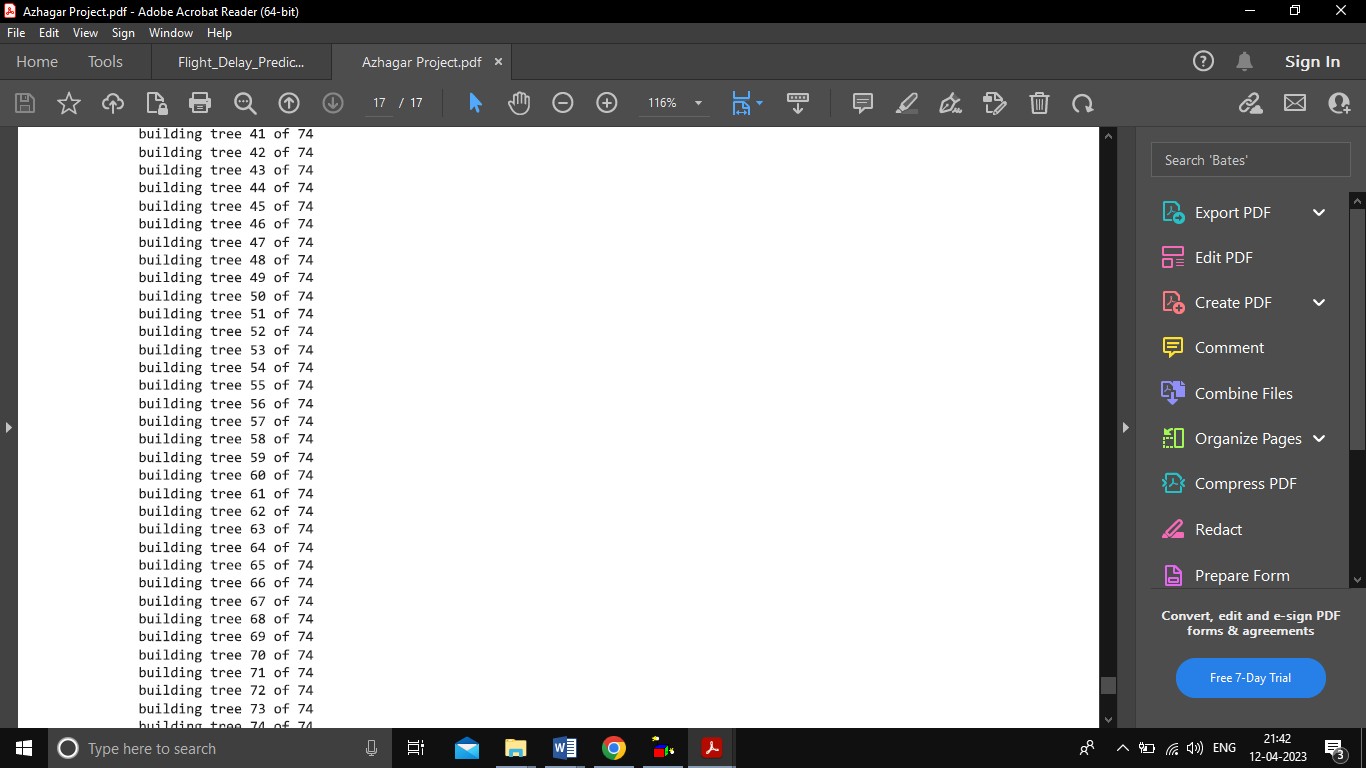












# ADVANTGES & DISADVANTAGES

* + Due to the stochastic nature of delays, this research investigates the qualitative prediction of airline delays to implement necessary changes and provide better customer experience.
  + Finding an accuracy of flight delay is less.
  + It does not have required parameters for finding flight delay.
  + Carriers attribute flight delays to several causes such as bad weather conditions, airport congestion, airspace congestion, and use of smaller aircraft by airlines. These delays and cancellations tarnish the airlines' reputation, often resulting in loss of demand by passengers.

# APPLICATION

* Flight Name
* Flight Scheduled Time
* Scheduled Date
* Source & Destination
* Passenger Details
* Speed
* Distance
* Running Status
* Delay Status Updating

# CONCLUSION

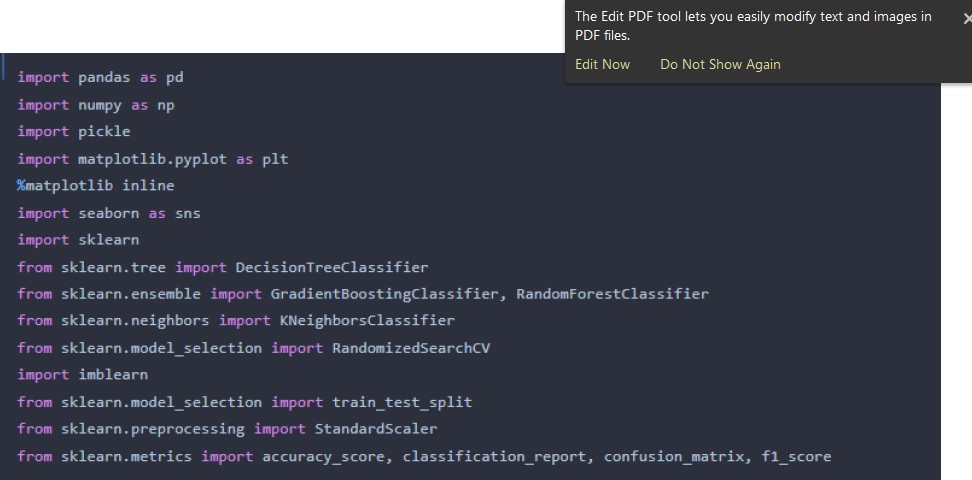
Predicting flight delays is on interesting research topic and required many attentions these years. Majority of research have tried to develop and expand their models in order to increase the precision and accuracy of predicting flight delays. Since the issue of flights being on-time is very important, flight delay prediction models must have high precision and accuracy. Based on the analysis of their results, it is evident that the integration of multidimensional heterogeneous data, combined with the application of different techniques for feature selection and regression can provide promising tools for inference in the cancer domain. Regardless of the type of prediction task at hand; regression or classification. It has become the state-of-the-art machine learning algorithm to deal with structured data. Compare to all algorithms MLP algorithm gives high accuracy that is 82%.

# FUTURE SCOP

The industry is expected to finally reach a profit in 2023 – the first time since the Covid- 19 pandemic – but unpredictable factors such as oil costs and ever-increasingly extreme weather patterns will continue to create more challenges for airlines.

# APPENDIX

A. Sourec Code

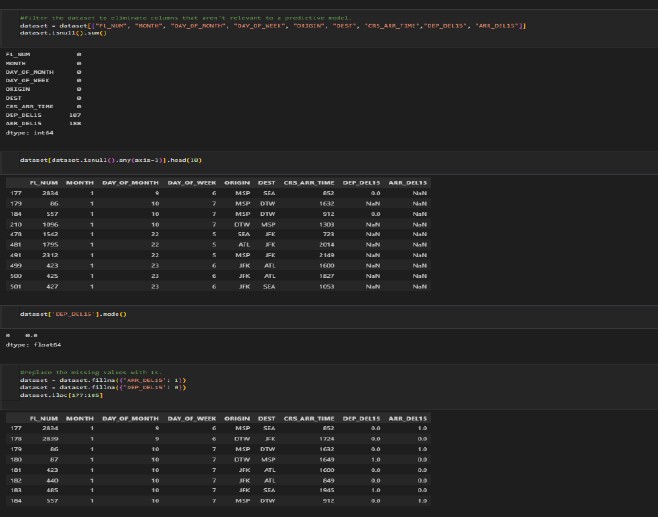


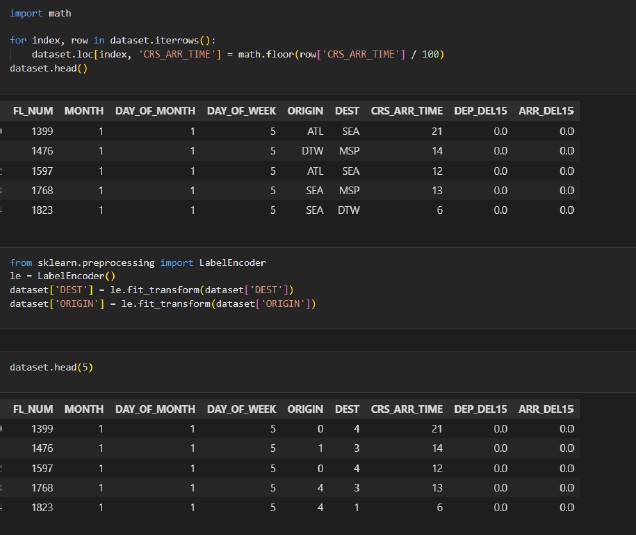
dataset= pd.read\_csv(“flightdata.csv”)

dataset.head() dataset.info()

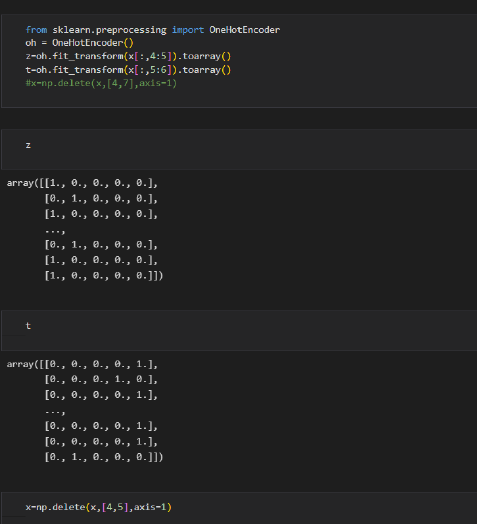
dataset = dataset.drop(‘Unnamed: 25’, axis=1

dataset.isnull().sum()





Dataset[‘ORIGIN’].unique()



1. **Data Analysis** Flight\_data.describe() sns.distplot(flight\_data.MONTH)

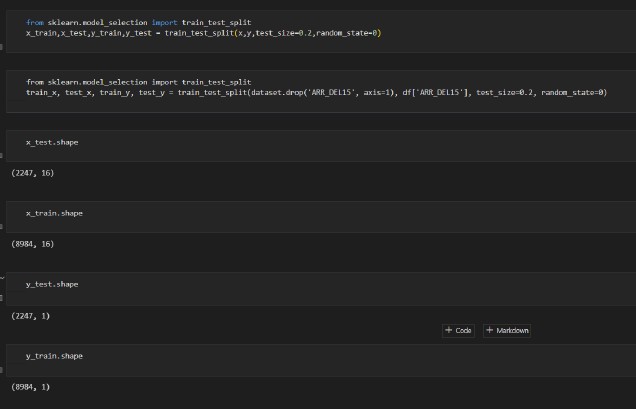
sns.scatterplot(x=’ARR\_DELAY’,y=’ARR\_DEL15’,data=flight\_data) sns.catplot(x=”ARR\_DEL”,y=”ARR\_DELAY”,kind=’bar’,data=flight\_data) sns.heatmap(dataset.corr())

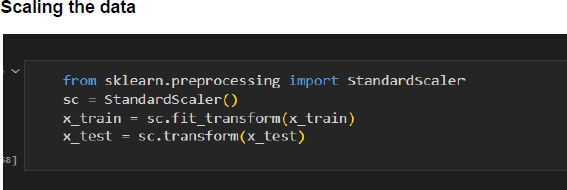
dataset = pd.get\_dummies(dataset, columns=[‘ORINGIN’, ‘DEST’])

dataset.head()

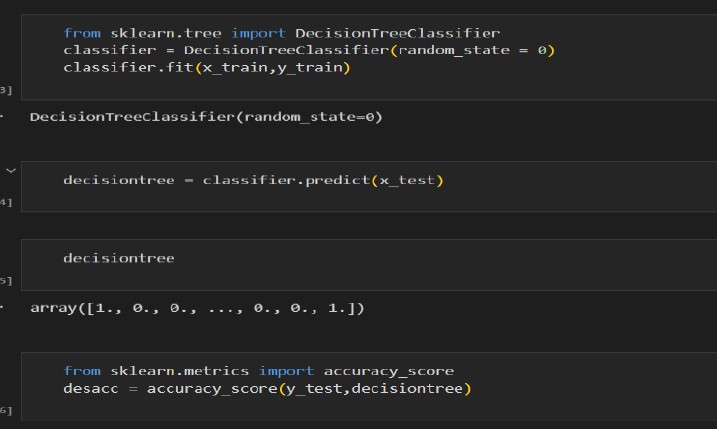
x = dataset.iloc[:, 0:8].values

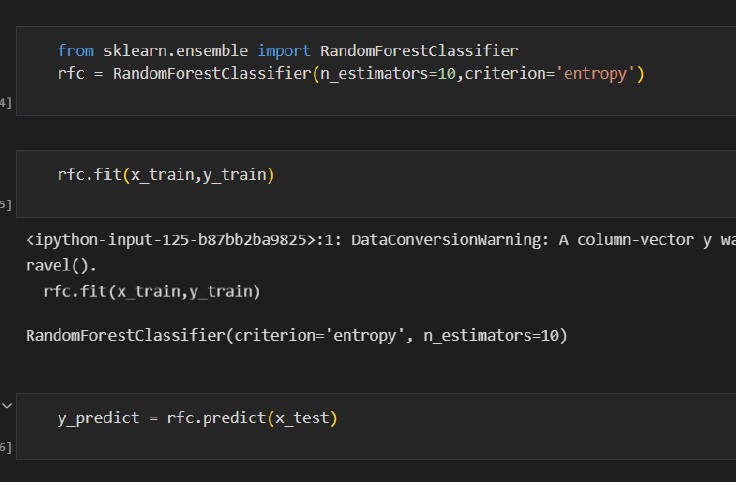
y = = dataset.iloc[:, 8:9].values

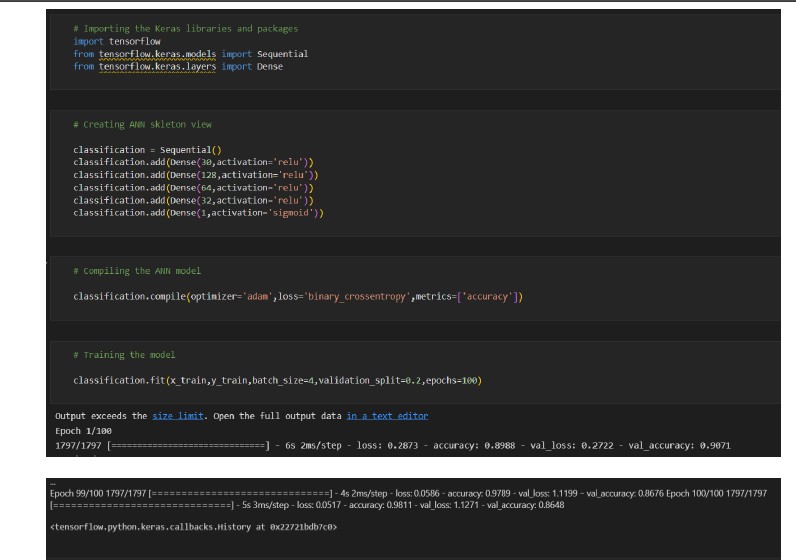


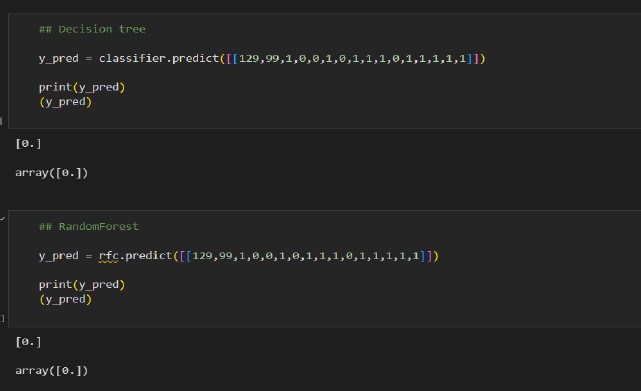


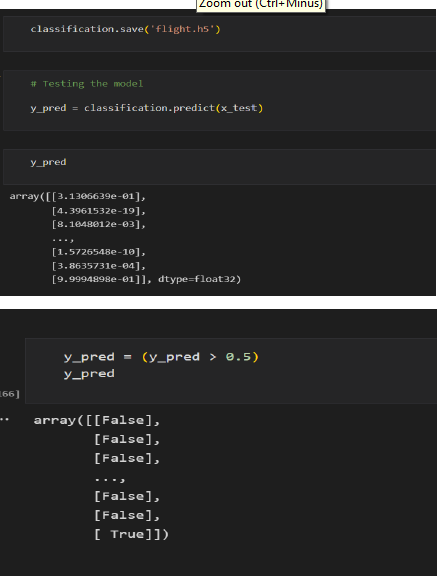
## Model Building

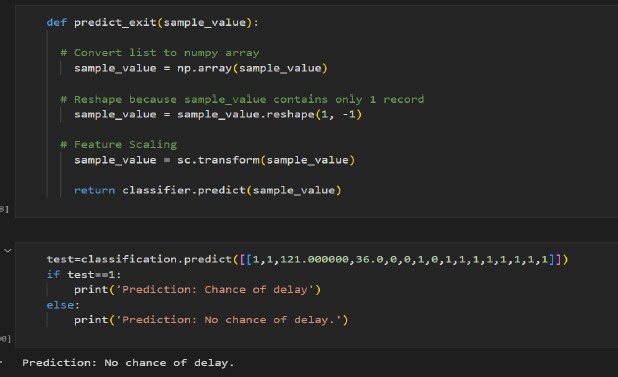
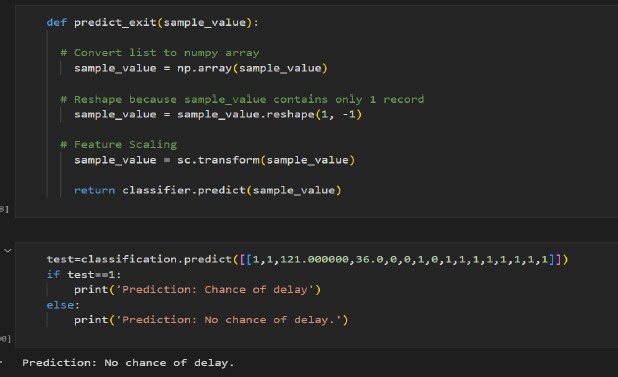










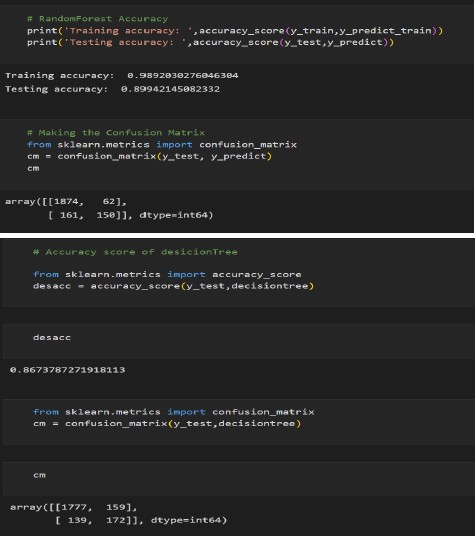


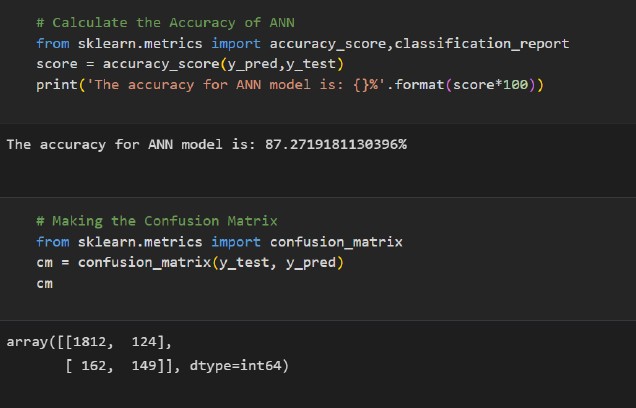
1. **Performance Testing & Hyperparameter Tuning**

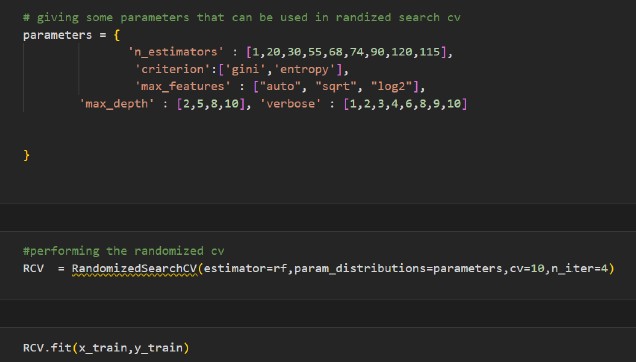
from sklearn import model\_selection

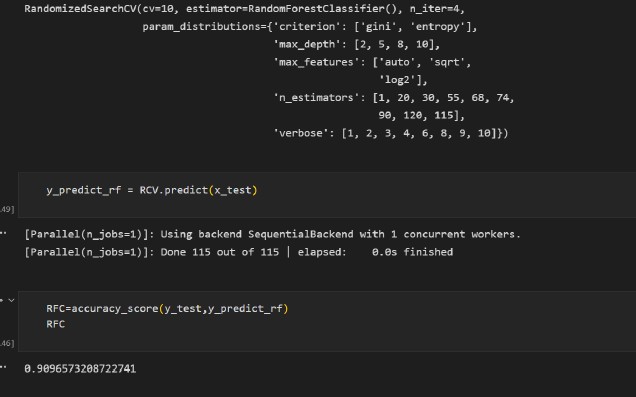
from sklearn.neural\_network import MLPClassifier











## Model Deployment

import pickle

pickle.dump(RCV,open(‘flight.pkl’,’wb’))

# importing the necessary dependencies

from flask import Flask, request, render\_template import numpy as np

import pandas as pd import pickle import os

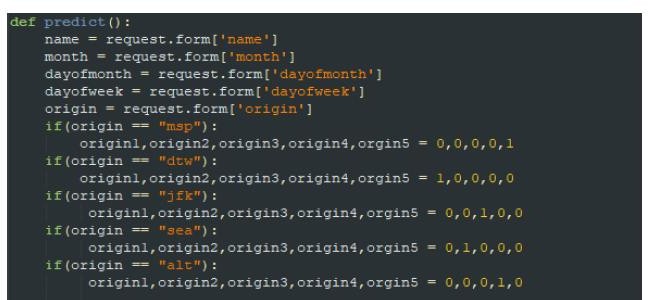
model = pickle.load(‘flight.pkl’,’rb’)

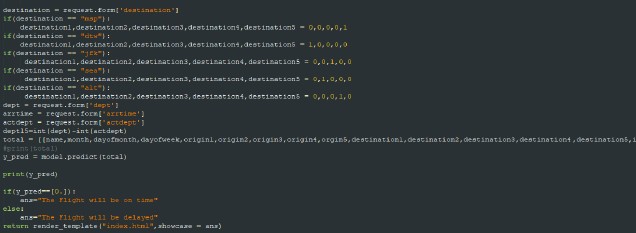
app = Flask( name )#initializing the app

@app.route(‘/’)

def home () :

return render\_template(“index.html”) @app.route(‘/prediction’,methods =[‘POST’])





If name == ‘ main ’:

app.run(debug = True)

