FLIGHT DELAY PREDICTION

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING(AI&ML)

Submitted By

GUNDA VENKATA SAI 21UK1A6664

THUMULA VYSHNAVI 21UK1A6612

BODDULA MANOJ 22UK5A6605

SAMALA SANTHOSH 21UK1A6638

Under the guidance of

N. RAJESH

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERINGVAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) – 506005

DEPARTMENT OF
COMPUTER SCIENCE AND ENGINEERING(AI&ML)

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION
INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled “FLIGHT DELAY PREDICTION” is being submitted by GUNDA VENKATA SAI(21UK1A6664), THUMULA VYSHNAVI (21UK1A6612), BODDULA MANOJ(22UK5A6605), SAMALA SANTHOSH(21UK1A6638). in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023- 2024.

Project Guide:	**HOD:**
N. RAJESH	**Dr. K. Sharmila**
(Assistant Professor)	(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr.P.PRASAD RAO,** Principal, Vaagdevi Engineering

College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.K.SHARMILA**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited,for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **N.RAJESH ,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

GUNDA VENKATA SAI(21UK1A6664), THUMULA VYSHNAVI (21UK1A6612), BODDULA MANOJ(22UK5A6605), SAMALA SANTHOSH(21UK1A6638).

ABSTRACT

Flight delays are a significant issue in the aviation industry, affecting both airlines and passengers by causing inconvenience, economic losses, and logistical challenges. This research focuses on predicting flight delays using a data-driven approach, integrating historical flight data, weather conditions, air traffic information, and other relevant variables. We employ advanced machine learning techniques, including regression models, decision trees, and neural networks, to analyze and predict delays. Our study highlights the importance of feature selection, data preprocessing, and model tuning in enhancing prediction accuracy.

Experimental results demonstrate that our proposed models can effectively predict delays with high accuracy, providing valuable insights for airlines to optimize their operations and improve passenger satisfaction. Future work will explore the incorporation of real-time data and the development of a robust predictive framework that can be deployed in operational environments.

TABLE OFCONTENTS:-

1.INTRODUCTION

1.1.OVERVIEW

The reliability and punctuality of flight schedules are crucial for the efficient operation of the aviation industry. Flight delays, however, remain a persistent challenge, impacting airlines, passengers, and broader economic activities. Delays can arise from various factors, including adverse weather conditions, air traffic congestion, technical issues, and operational inefficiencies. The unpredictability and cascading effects of delays can lead to substantial economic losses and diminished passenger satisfaction.

The advancement of data analytics and machine learning has opened new avenues for enhancing the accuracy of flight delay predictions. Historical flight data, combined with real-time information on weather, air traffic, and other operational variables, can be leveraged to develop sophisticated predictive models. These models can identify patterns and correlations that are not apparent through traditional analytical methods.

This research aims to explore various machine learning techniques for predicting flight delays. We examine the effectiveness of different models, such as regression algorithms, decision trees, and neural networks, in forecasting delays. The study also emphasizes the importance of feature selection, data preprocessing, and model evaluation in achieving reliable predictions.

By providing a comprehensive analysis of flight delay prediction methodologies, this research seeks to contribute to the development of robust tools that can be integrated into airline and airport operations. Ultimately, the goal is to enhance the efficiency and reliability of the aviation sector, thereby improving the overall travel experience for passengers.

1.2.PURPOSE

The primary purpose of flight delay prediction is to enhance the operational efficiency and service quality of the aviation industry. By accurately forecasting potential delays, stakeholders can proactively manage and mitigate the adverse effects associated with flight schedule disruptions. The specific objectives include:

1. **Improving Operational Efficiency:**
    - **Resource Allocation:** Accurate delay predictions enable airlines and airports to better allocate resources, such as crew, ground staff, and gates, ensuring smoother operations.
    - **Schedule Optimization:** Airlines can optimize flight schedules by adjusting departure and arrival times based on predicted delays, minimizing turnaround times and reducing the risk of subsequent delays.
2. **Enhancing Passenger Experience:**
    - **Timely Communication:** Predicting delays allows airlines to provide timely information to passengers, reducing uncertainty and frustration.
3. **Reducing Economic Impact:**
    - **Cost Savings:** By anticipating delays, airlines can implement cost-saving measures, such as fuel management strategies and efficient crew scheduling, thereby reducing operational costs.
4. **Optimizing Air Traffic Management:**
    - **Air Traffic Flow:** Accurate delay predictions assist air traffic controllers in managing airspace more effectively, reducing congestion and enhancing the overall flow of air traffic.
5. **Supporting Strategic Planning:**
    - **Long-term Insights:** Historical and predictive data analysis provides valuable insights for long-term strategic planning, helping airlines and airports identify recurring issues and develop strategies to address them.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

Despite the advancements in data analytics and machine learning, flight delay prediction still faces several significant challenges. These existing problems can hinder the accuracy and reliability of predictive models, impacting their effectiveness in real-world applications. Key challenges include:

1. **Data Quality and Availability:**
   - **Incomplete Data:** Missing or incomplete historical data can lead to biased or inaccurate models. Inconsistent data reporting from various airlines and airports further complicates the issue.
   - **Data Granularity:** The granularity of available data can vary significantly, with some datasets lacking detailed information on factors like specific weather conditions or precise air traffic details.
2. **Complexity of Influencing Factors:**
   - **Multifactorial Nature:** Flight delays are influenced by a myriad of factors, including weather, air traffic control decisions, technical issues, and operational constraints. Capturing the interplay between these factors is complex.
   - **Dynamic Interactions:** The dynamic nature of factors such as changing weather conditions and real-time air traffic makes it challenging to predict delays accurately.
3. **Model Limitations:**
   - **Overfitting and Underfitting:** Predictive models can suffer from overfitting, where they perform well on training data but poorly on new data, or underfitting, where they fail to capture the underlying patterns in the data.
   - **Model Interpretability:** Complex machine learning models, such as neural networks, often lack interpretability, making it difficult to understand how predictions are made and to trust their outputs.

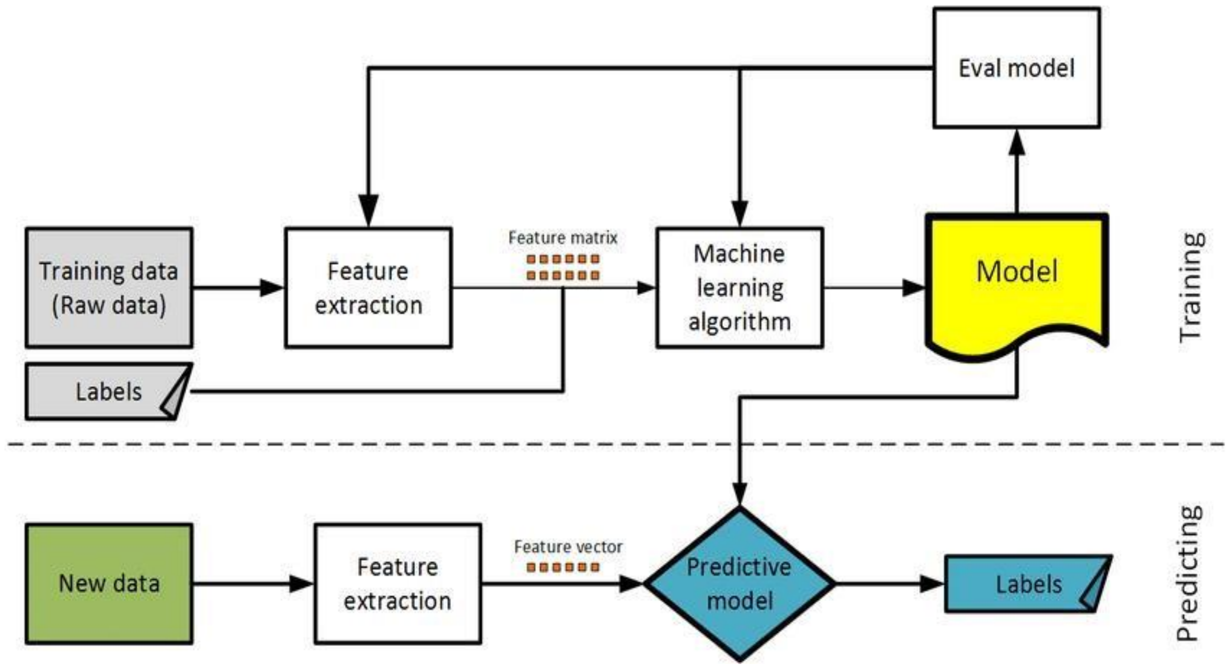
2.2 PROPOSED SOLLUTION

To address the existing challenges in flight delay prediction, a multifaceted approach combining advanced data analytics, machine learning techniques, and collaborative efforts is essential. The proposed solutions include:

1. **Enhanced Data Collection and Management:**
    - **Comprehensive Data Integration:** Integrate diverse data sources, including weather forecasts, historical flight data, air traffic control information, and airport operational data, to create a comprehensive dataset.
2. **Advanced Machine Learning Techniques:**
    - **Hybrid Models:** Develop hybrid models that combine different machine learning techniques, such as regression, decision trees, and neural networks, to capture the complex relationships between various factors influencing flight delays.
3. **Feature Engineering and Selection:**
    - **Domain-Specific Features:** Identify and engineer domain-specific features that are known to influence flight delays, such as wind speed, runway availability, and crew schedules.
4. **Improved Model Interpretability:**
    - **Explainable AI:** Incorporate explainable AI techniques to make complex models more interpretable, allowing stakeholders to understand the reasoning behind predictions and build trust in the models.
5. **Real-time Prediction Capabilities:**
    - **Streaming Data Processing:** Implement streaming data processing frameworks to handle real-time data inputs and provide timely predictions, minimizing latency and ensuring up-to-date information.

3.THEORITICALANALYSIS

3.1. BLOCK DIAGRAM



3.2. SOFTWARE DESIGNING

The following is the Software required to complete this project:

➢ **Jupyter Notebook**: Jupyter Notebook will serve as the development and execution environment for your predictive modeling, data preprocessing, and model training tasks. It provides a cloud-based Jupyter Notebook environment with access to Python libraries and hardware acceleration.

➢ **Dataset (CSV File)**: The dataset in CSV format is essential for training and testing your predictive model. It should include historical scheduled arrivaltime,scheduled depaturetime,origin, delaytime,weather information, and other relevant features.

➢ **Data Preprocessing Tools**: Python libraries like NumPy, Pandas, and Scikit-learn will be used to preprocess the dataset. This includes handling missing data, feature scaling, and data cleaning.

- ➢ **Feature Selection/Drop**: Feature selection or dropping unnecessary features from the dataset can be done using Scikit-learn or custom Python code to enhance the model's efficiency.
- ➢ **Model Training Tools**: Machine learning libraries such as Scikit-learn, logistic regression will be used to develop, train, and fine-tune the predictive model. Regression or classification models can be considered, depending on the nature of flight delay prediction.
- ➢ **Model Accuracy Evaluation**: After model training, accuracy and performance evaluation tools, such as Scikit-learn metrics or custom validation scripts, will assess the model's predictive capabilities. You'll measure the model's ability to predict flight delay categories based on historical data.
- ➢ **UI Based on Flask Environment**: Flask, a Python web framework, will be used to develop the user interface (UI) for the system. The Flask application will provide a user-friendly platform for users to input location data or view flight delay predictions, health information, and recommended precautions.
- ➢ Jupyter Notebook will be the central hub for model development and training, while Flask will facilitate user interaction and data presentation. The dataset, along with data preprocessing, will ensure the quality of the training data, and feature selection will optimize the model. Finally, model accuracy evaluation will confirm the system's predictive capabilities, allowing users to rely on the Flight delay predictions and associated health information.

4.EXPERIMENTAL INVESTIGATION

In this project, we have used Flight Delay prediction Dataset. This dataset is a csv file consisting of labelled data and having the following columns-

1. **Flight Number:**It displays the specific flight number.

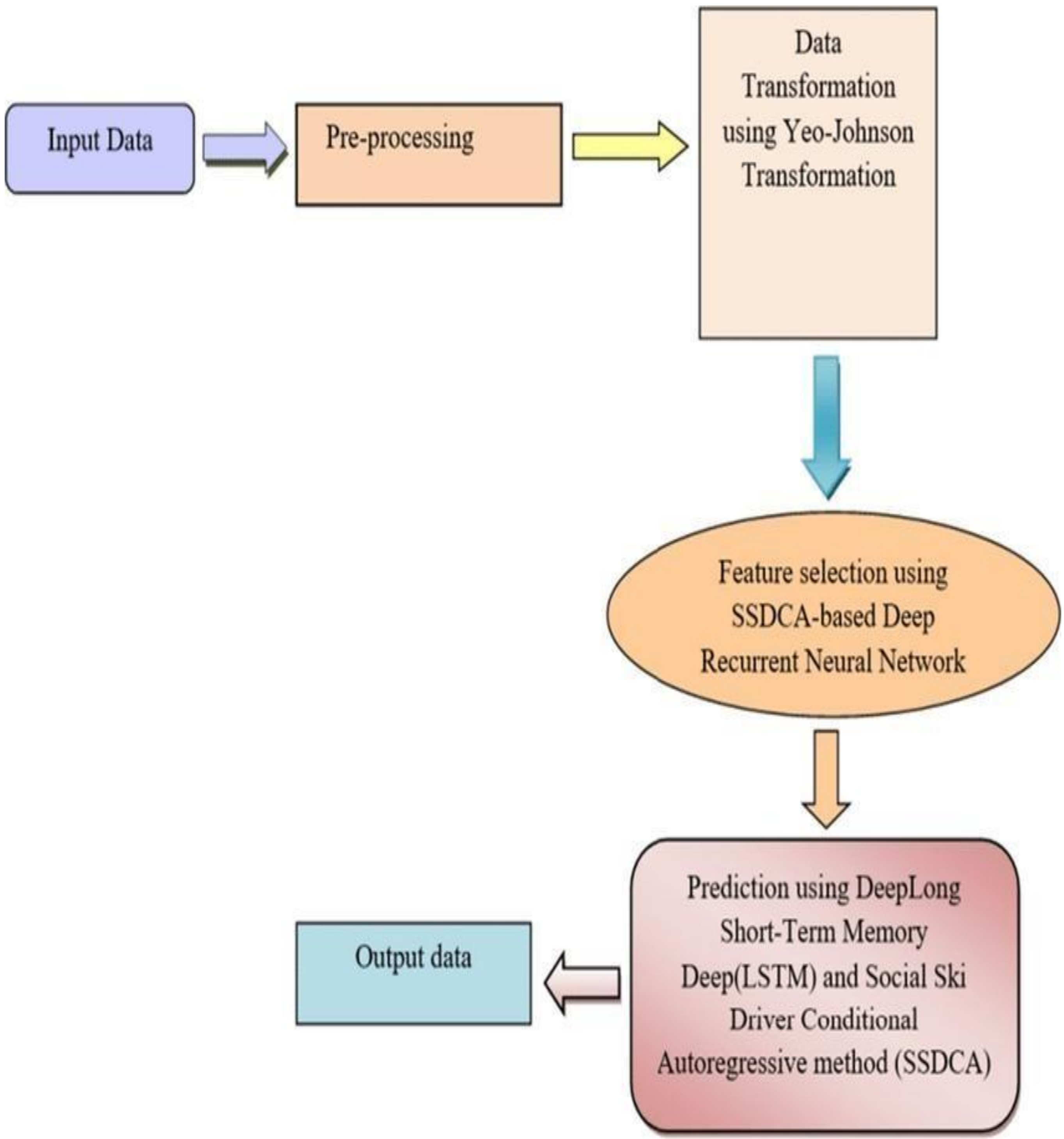
2. **Month**: The month of December often sees the most flight delays due to winter weather and holiday travel.
3. **Day of month**: he 27th of December often experiences significant flight delays due to post-Christmas
4. **Day of week**: Fridays typically have the most flight delays due to increased weekend travel.
5. **Origin**: The origin (departure) airport plays a significant role in flight delays prediction due to local weather and traffic conditions.
6. **Destination**:The destination airport is crucial in flight delays prediction due to its weather conditions, traffic congestion, and operational efficiency.
7. **Departure delay**: Departure delays significantly impact the likelihood of subsequent flight delays throughout the journey.
8. **Scheduled ArrivalTime**: Scheduled arrival time affects flight delays prediction as peak hours often see more congestion and delays.
9. **Arrival delay**: Arrival delays indicate potential bottlenecks at the destination, impacting the prediction of subsequent flight delays.

For the dataset we selected, it consists of more than the columns we want to predict it . So, we have chosen the feature drop it contains the columns that we are going to predict the Flight delays.

➢ Feature drop means it drops the columns that we don’t want in our dataset.

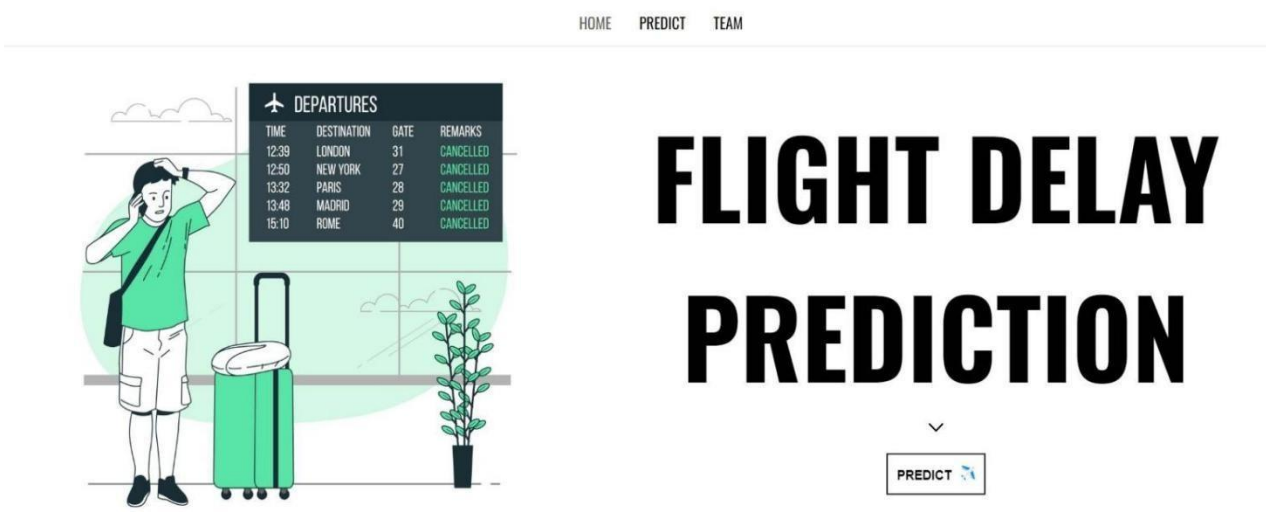
➢ Feature_drop=[`flight-no`,`day-of-month`,`day-of-week`,`origin`,`destination`, `departure delay`,`scheduled arrivaltime`,`arrivaldelay`]

5 . FLOW CHART

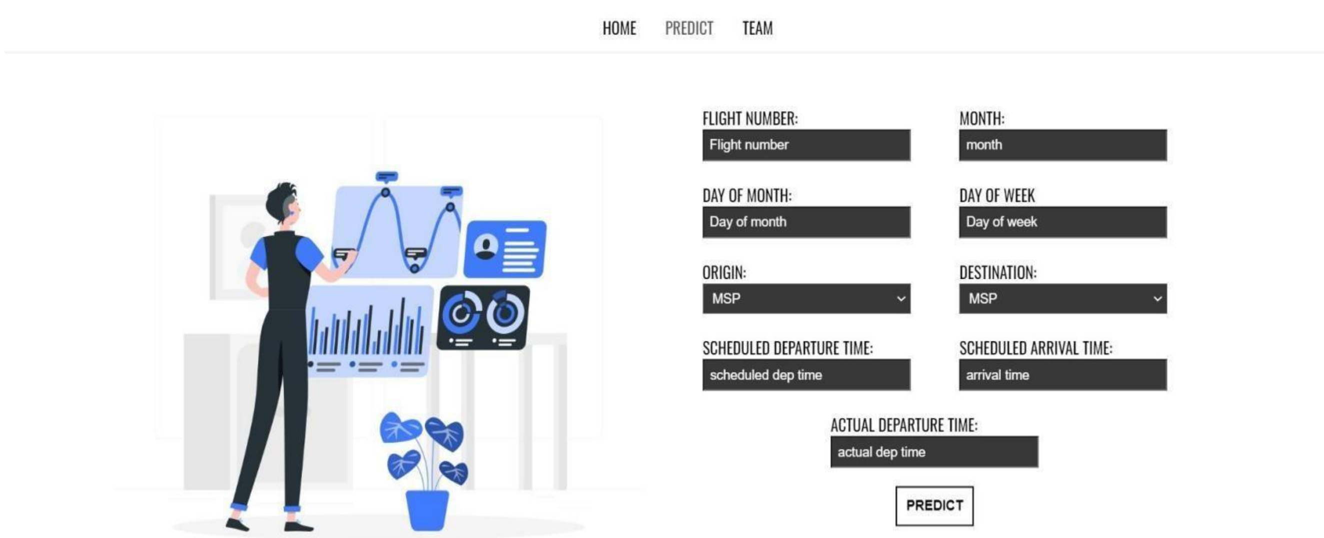


6.RESULT

HOME PAGE:



PREDICTION PAGE:



RESULT PAGE:

7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

1. **Improved Customer Satisfaction**:

   - **Proactive Communication**: Passengers can be informed in advance about potential delays, allowing them to adjust their plans accordingly.
   - **Enhanced Planning**: Travelers can make informed decisions about their journeys, including connections and accommodations.

2. **Operational Efficiency for Airlines**:

   - **Resource Optimization**: Airlines can better manage gate assignments, crew schedules, and aircraft utilization, leading to more efficient operations.
   - **Cost Savings**: Predicting delays can help reduce costs associated with missed connections, passenger compensations, and last-minute operational adjustments.

3. **Safety Enhancements**:

- **Risk Mitigation**: By anticipating delays, airlines can implement measures to ensure safety and avoid rushed or compromised procedures.

4. **Competitive Advantage**:

- **Reputation Management**: Airlines that reliably predict and manage delays can enhance their reputation and attract more customers.
- **Market Differentiation**: Offering predictive delay services can be a unique selling point.

5. **Improved Airport Operations**:

- **Traffic Management**: Airports can better manage runway and gate usage, reducing congestion and enhancing overall efficiency.

DISADVANTAGES:

1. **Data Complexity and Volume**:

- **Massive Data Requirements**: Predicting flight delays requires a vast amount of data, including weather conditions, air traffic, historical flight data, and real-time aircraft positions. Managing and processing this data can be complex and resource-intensive.

2. **Model Limitations**:

- **Complex Modeling**: Flight delay prediction involves numerous variables and interactions, making it challenging to create accurate predictive models. These models must account for dynamic factors like sudden weather changes or unexpected air traffic issues.

3. **Unpredictable Factors**:

- **Weather Variability**: Weather is a significant factor in flight delays, but it is inherently unpredictable. Sudden storms or changes in weather conditions can cause delays that models might not anticipate.

4.**Operational Challenges**:

- **Integration with Airline Operations**: Implementing predictive models within airline operations can be challenging. Airlines need to integrate predictions with their scheduling, crew management, and passenger communication systems.

5. **Economic and Logistic Constraints**:

 Cost: Developing, maintaining, and updating predictive models can be costly. This includes expenses for data acquisition, storage, processing, and the technical infrastructure required.

8.APPLICATIONS

1. **Operational Efficiency**: Airlines can optimize crew scheduling, gate assignments, and ground operations based on predicted delays, leading to cost savings and improved efficiency.

2. **Passenger Satisfaction**: Passengers can be informed in advance about potential delays, allowing them to adjust their travel plans accordingly or providing them with alternative options.

3. **Resource Management**: Airports can better manage resources such as runway usage, fueling, and baggage handling by anticipating delays and adjusting their operations accordingly.
4. **Air Traffic Control**: Predicting delays helps air traffic controllers manage air traffic flow more effectively, reducing congestion and improving overall safety.

5. **Financial Planning**: Airlines can better manage their financial planning and minimize losses associated with delays by predicting when and where delays are likely to occur.

9.CONCLUSION

In conclusion, flight delay prediction using machine learning offers substantial benefits across various facets of the aviation industry. By leveraging historical data, real-time information, and advanced algorithms, machine learning models can accurately forecast potential delays. This capability enables airlines to enhance operational efficiency, improve passenger satisfaction, optimize resource allocation, and support better decision-making for air traffic management. Moreover, continuous analysis of delay patterns provides valuable insights for long-term planning and infrastructure development, ultimately contributing to safer and more reliable air travel experiences. As technology evolves and datasets expand, the efficacy of machine learning in predicting flight delays will likely continue to advance, offering even greater advantages to stakeholders in the aviation ecosystem.

10.FUTURE SCOPE

The future scope of flight delay prediction using machine learning (ML) is promising, with several advancements and opportunities on the horizon:

1. **Enhanced Accuracy**: As ML algorithms continue to evolve, they will become more adept at analyzing complex data patterns and incorporating real-time information. This will lead to even more accurate predictions of flight delays.
2. **Integration of Big Data**: Integration with big data sources, such as social media, airport operations data, and aircraft sensor data, will enrich predictive models. This holistic approach will provide a comprehensive view of factors influencing delays.

3. **Real-time Decision Support**: ML-powered systems will offer real-time decision support to airlines, airports, and air traffic controllers. This capability will enable proactive management of delays and optimize operations dynamically.

4. **Personalized Passenger Information**: Passengers will receive personalized notifications and alternative travel options in case of predicted delays, improving their travel experience and satisfaction.

5. **Predictive Maintenance**: ML can also contribute to predicting aircraft maintenance needs more accurately, reducing unplanned maintenance-related delays.

11.BIBILOGRAPHY

- **Books**:

- Belobaba, P. (2009). *The Global Airline Industry*. John Wiley & Sons.
- Wilson, R. (2011). *Airport Operations*. McGraw-Hill Education.

- **Research Papers**:

- Hansman, R. J., & Zeghal, K. (1999). Predicting airport delays. *Transportation Science, 33*(3), 240-249.
- Ball, M. O., Barnhart, C., & Weatherford, L. R. (2002). Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science, 36*(1), 3-16.

- **Journal Articles**:
- Lei, H., Wu, S., & Fu, X. (2017). A review on the flight delay prediction problem. *Journal of Air Transport Management, 62*, 263-276.
- Dong, Y., Jin, P., & Zhang, Y. (2019). Flight delay prediction with machine learning methods: A review from 2015 to 2018. *Journal of Air Transport Management, 79*, 101686.

- **Conference Proceedings**:

- Borenstein, N., Herer, Y., & Kaplan, S. (1994). Predicting delay propagation in the national airspace system. In *Proceedings of the 33rd IEEE Conference on Decision and Control* (Vol. 2, pp. 1790-1795). IEEE.

- **Online Resources**:

- FAA. (2020). Airport Delays. Federal Aviation Administration. Retrieved from https://www.faa.gov/airports/airport_safety/airport_delays/

12.APPENDIX

Model building :

1)Dataset
2)Jupyter Notebook and VS code Application Building
  1. HTML file (Home file, Predict file)
  2.CSS file
  3.Models in pickle format

SOURCE CODE:

HOME.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <link rel="stylesheet" type="text/css" href="../static/style/style.css">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Document</title>
</head>
<body>
   <nav>
      <div>
         <ul>
            <a href='{{url_for("home")}}' class="active"><li>HOME</li></a>
            <a href='{{url_for("predict")}}'><li>PREDICT</li></a>
```

```
            <a href='{{url_for("team")}}'><li>TEAM</li></a>
          </ul>
        </nav>
    </div>
    <div class="home-container">
        <div>
         <img src="../static/images/4944426.jpg" width="500px">
        </div>
        <div>
          <section>
            <div>
               <h1 class="title title1">FLIGHT DELAY</h1>
               <h3 class="title title2">PREDICTION</h3>
            </div>
          </section>
          <div class="btn-div">
            <svg  xmlns="http://www.w3.org/2000/svg"  width="28"  height="28"  viewBox="0  0  24  24"><path
fill="none" stroke="currentColor" stroke-linecap="round" stroke-linejoin="round" stroke-width="1.5" d="m6 9l6
6l6-6"/></svg>
            <a href='{{url_for("predict")}}'><button class="btn" >PREDICT<img src="../static/images/airplane.png"
width="15px" style="display:inline-block;margin-left:10px;margin-top:1px;"></button></a> </div><br/><br/>
        </div>
    </div><br/><br/><br/><br/>
   <hr/>
    <section class="content-sec">
         <div class="img-div">
            <img src="../static/images/flight-delays1.webp" alt="delay-img" width="500px" height="350px">
         </div>
         <div class="content-div">
          Flight delays can cause significant disruptions and inconveniences for travelers, leading to a host of
problems. These issues include schedule disruptions, missed connections, and the resulting frustration and
anxiety. Passengers often endure extended waiting times at the airport, facing discomfort and limited access to
amenities. The financial burden of unexpected expenses for meals, accommodations, or alternative
transportation adds to the stress. Health and comfort concerns arise from prolonged periods of sitting and
waiting, while the uncertainty about future travel plans, including hotel reservations and scheduled activities, can
be overwhelming.
        </div>
    </section>
    <br/><br/>
    <hr/>
    <section class="content-sec2">
        <div class="content-div2">
```

Understanding the severity of these challenges, I was inspired to create a flight delay prediction system using machine learning. By analyzing vast amounts of historical flight data, weather patterns, air traffic control information, and other relevant factors, this system can accurately predict potential delays. This advanced technology allows airlines and passengers to better prepare for and mitigate the impact of delays.
With this flight delay prediction system, passengers can receive real-time updates and make informed decisions about their travel plans. Airlines can optimize their operations, improve scheduling, and enhance overall customer satisfaction. By reducing the uncertainty and stress associated with flight delays, this innovative solution aims to improve the travel experience for everyone involved.

```
        </div>
        <div class="img-div2">
            <img src="../static/images/analytics.gif" width="400px">
        </div>
    </section>
    <hr/>
  <script src="../static/flight.js"></script>
</body>
</html>
```

PREDICT.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <link rel="stylesheet" type="text/css" href="../static/style/style.css">
    <link rel="stylesheet" type="text/css" href="../static/style/predict.css">
    <title>Document</title>
</head>
<body>
    <nav>
        <div>
            <ul>
                <a href='{{url_for("home")}}'><li>HOME</li></a>
                <a href='{{url_for("predict")}}' class="active"><li>PREDICT</li></a>
                <a href='{{url_for("team")}}'><li>TEAM</li></a>
            </ul>
        </div>
    </nav>
    <div class="container">
        <div>
            <img src="../static/images/Data analysis-rafiki.png" width="550px">
        </div>
        <form action='{{url_for("result")}}' method="post">
            <div>
```

```
    <div class="inputs">
        <label>FLIGHT NUMBER:</label><br/>
        <input type="number" required name="fltnum" placeholder="Flight number">
    </div>
    <div class="inputs">
        <label>MONTH:</label><br/>
        <input type="number" min="1" max="12" required name="month" placeholder="month">
    </div>
</div>
<div>
    <div class="inputs">
        <label>DAY OF MONTH:</label><br/>
        <input type="number"required min="1" max="30" name="dayofmonth" placeholder="Day of month">
    </div>
    <div class="inputs">
     <label>DAY OF WEEK</label><br/>
     <input type="number" min="1" max="7" required name="dayofweek" placeholder="Day of week">
    </div>
</div>
<div>
    <div class="inputs">
        <label>ORIGIN:</label><br/>
        <select name="origin">
            <option value="msp">MSP</option>
            <option value="dtw">DTW</option>
            <option value="jfk">JFK</option>
            <option value="sea">SEA</option>
            <option value="atl">ALT</option>
        </select>
    </div>
    <div class="inputs">
        <label>DESTINATION:</label><br/>
        <select name="destination">
            <option value="msp">MSP</option>
            <option value="dtw">DTW</option>
            <option value="jfk">JFK</option>
            <option value="sea">SEA</option>
            <option value="atl">ALT</option>
        </select>
    </div>
</div>
<div>
```

```
            <div class="inputs">
              <label>SCHEDULED DEPARTURE TIME:</label><br/>
              <input type="number" min="1" name="dept" required placeholder="scheduled dep time">
            </div>
            <div class="inputs">
              <label>SCHEDULED ARRIVAL TIME:</label><br/>
              <input type="number" min="1" name="arrtime" required placeholder="arrival time">
            </div>
          </div>
          <div class="inputs">
            <label>ACTUAL DEPARTURE TIME:</label>
            <input type="number" name="acttime" min="1" required placeholder="actual dep time">
          </div>
          <div class="inputs">
            <button class="btn" type="submit">PREDICT</button> </div>
        </form>
      </div>
      <script src="../static/flight.js"></script>
    </body>
    </html>
```

2. CSS FILE:

STYLE.CSS

```
@font-face     {     font-family:myfont;     src:

  url("../Oswald-VariableFont_wght.ttf")

}

*{

  margin:0;

}

body::-webkit-scrollbar{

  display:none;

}
.home-container{ display:flex;

  justify-content:space-around;

  margin-bottom:70px;

 /* margin-top:100px; */
```

```
}
.home-container > div:nth-child(2){ margin-
   top:50px ;
}
nav{ height:50px; border-bottom:2px solid rgba(238,
   238, 238, 0.707); font-family:myfont; font-
   weight:400; position:sticky; background-
   color:white; top:0px; z-index:999;
} ul{ padding-top:14px;
list-style:none;
display:flex; gap:30px;
justify-content:center;
align-items:center;
} a{ text-
decoration:none;
color:black;
} a li{ transition:0.3s
ease;
}
a li:hover{ color:rgb(114,
  113, 113);
  transform:scale(0.97,0.97);
}
.active{ color:rgb(93,
   91, 91);
   position:relative;
}
```

```
section{ text-
   align:center; font-
   family:myfont;
}
.title1,.title2{ font-
   size:110px; color:rgb(1, 1,
   1); opacity:0;

}
.visible{ position:relative; animation:slide 1.5s cubic-bezier(0.175,
   0.885, 0.32, 1.275) forwards; animation-delay:0.5s;
}
@keyframes slide {
   from{
      top:100px;
   } to{
    top:0px;
    opacity:1;
   }

}
.btn-div{ display:flex;
  flex-direction:column;
  justify-content:center;
  align-items:center;
  gap:10px; opacity:0;
}
```

```
.btn{ padding:10px; font-weight: bold; border:2px solid black;
  background-color:rgba(29, 252, 51, 0); transition:.5s ease;
  animation:float 1s cubic-bezier( 0.33, 0.74, 0.48, 0.65 ) infinite alternate;
}
@keyframes float { from{
   transform:translatey(5px);
   }
   to{ transform:translatey(-
      5px); opacity:1;
   }
}
.btn:hover{ background-
   color:black; color:white;
   transform:scale(1.05,1.05);
}
.content-sec{ margin-
   top:60px;
   padding:50px;
   display:flex;
   gap:40px;
}
.img-div{ opacity:0;
   position:relative;
}
.img-visible{ position:relative;
   animation:imgslide 1.5s ease forwards;
   animation-delay:0.5s;
}
```

```
@keyframes imgslide {
  from{ left:-100px;
  opacity:0;
   } to{
   left:0px;
   opacity:1;
   }


}
.content-div{ font-
  size:20px; text-align:
  justify; opacity:0;
  width:650px; line-
  height:37px;
  position:relative;
}
.content-visible{ position:relative;
  animation:contentslide 1.5s ease forwards;
  animation-delay:0.5s;
}
@keyframes contentslide {
  from{ opacity:0; right:-
  100px;
   } to{
   right:0px;
   opacity:1;
   }
}
```

```
.content-sec2{ margin-
   top:20px; padding:50px;
   display:flex; justify-
   content:space-around;
}
.img-div2{
   opacity:0;
}
.content-div2{ margin-
   top:20px; font-
   size:20px; text-align:
   justify; width:650px;
   opacity:0; line-
   height:37px;
   position:relative;
}
.container{ display:flex;
   font-family: myfont;
   justify-content:center;
   align-items:flex-start;
   height:500px;
   overflow:hidden;
}
form{
  margin-top:50px;
  width:600px; display:flex;
  flex-direction:column; justify-
  content:space-around;
```

```
gap:20px; align-items:center;
border-radius:10px;
}
input{
padding:6px;
width:200px;
} select{
padding:6px;
width:215px;
}
form > div{
display:flex; justify-content:space-
around; gap:50px;
}
form >div:nth-child(5){ flex-
direction: column; gap:0px;
}
form > button{ padding:10px;
background-color:transparent;
border:2px solid black;
transition:0.4s ease; font-
weight:bold; animation:float
1s cubic-bezier( 0.33, 0.74,
0.48, 0.65 ) infinite alternate;
}
form > button:hover{ background-
color:black; color:white;
transform:scale(1.06,1.06);
```

```
}
.predictbg{
   position:absolute;
   width:98vw;
   height:99vh; z-index:-
   99;
} input,select{ background-
color:rgba(0, 0, 0, 0.78); backdrop-
filter:blur(10px); color:white;
}
input::placeholder{
   color:white;
}
.result-container{
   display:flex; font-
   family:myfont; width:
   100vw; height:100vh;
   flex-direction: column;
   justify-content:center;
   align-items:center;
}
.team-container{
  /* background-color: rgba(0, 0, 0, 0.734); */
  color:rgb(1, 1, 1); font-family:myfont;
  display:flex; justify-content:center; align-
  items:center; gap:10px; margin-top:100px;
}
```

```
.div-1{ width:210px;
 height:200px; background-
 color:rgb(0, 0, 0); border:2px
 solid white; position:relative;
 transition:1s ease;
}
.div-1 > div:nth-child(1){
   height:200px;
   width:210px;
   display:flex; justify-
   content:center; flex-
   direction: column; align-
   items:center;
   position:absolute;
   color:rgb(0, 0, 0);
   background-color: white;
   transition:1s ease;
}
.div-1 > div:nth-child(2){
   display:flex; justify-
   content:center; align-
   items:center; position:
   absolute; top:0px;
   background-color:black;
   height:200px;
   width:210px;
   color:white;
   transition:1s ease;
```

```
}
.div-1:hover div:nth-child(2){
   opacity:0;
}
.inputs{ position:relative; opacity:0;
  animation:slide3 1s ease forwards;
}
@keyframes slide3 {
   from{ left:400px;
   }
   to{ left:0px;
   opacity:1;
   }
}
.loader { width: 100px; height: 3px;
   background-color: rgb(15, 15, 15);
   border-radius:   20px;   overflow:
   hidden;
  }

  .child { width: 60px; height: 3px;
   background-color: rgb(107, 27, 255);
   border-radius: 20px; z-index: 0;
   margin-left: -60px; animation: go 1s
   0s infinite;
  }
```

```
@keyframes go {
  from { margin-left: -
  100px; width: 80px;
  }
  to {
    width: 30px; margin-
    left: 110px;
  }
}

ul { list-style:
  none;
}

.text { width: 100px; height:
  30px; background-color:
  transparent; margin-top: 20px;
  text-align: center;
}

.text::before   {   content:
  "Loading";  color:black;  font-
  family:myfont;    animation:
  text 1s 0s infinite;
}

@keyframes text {
  0% {
```

```
    content: "Predicting";
   }
   30% {
    content: "Predicting.";
   }

   60% {
    content: "Predicting..";
   }

   100% {
    content: "Predicting...";
   }
  }
.result-main-con{
  height:500px; display:flex;
  justify-content:space-around;
  background-color:white;
}
.result-main-con >div:nth-child(1){ margin-
  left:50px;
}
```

APP.PY

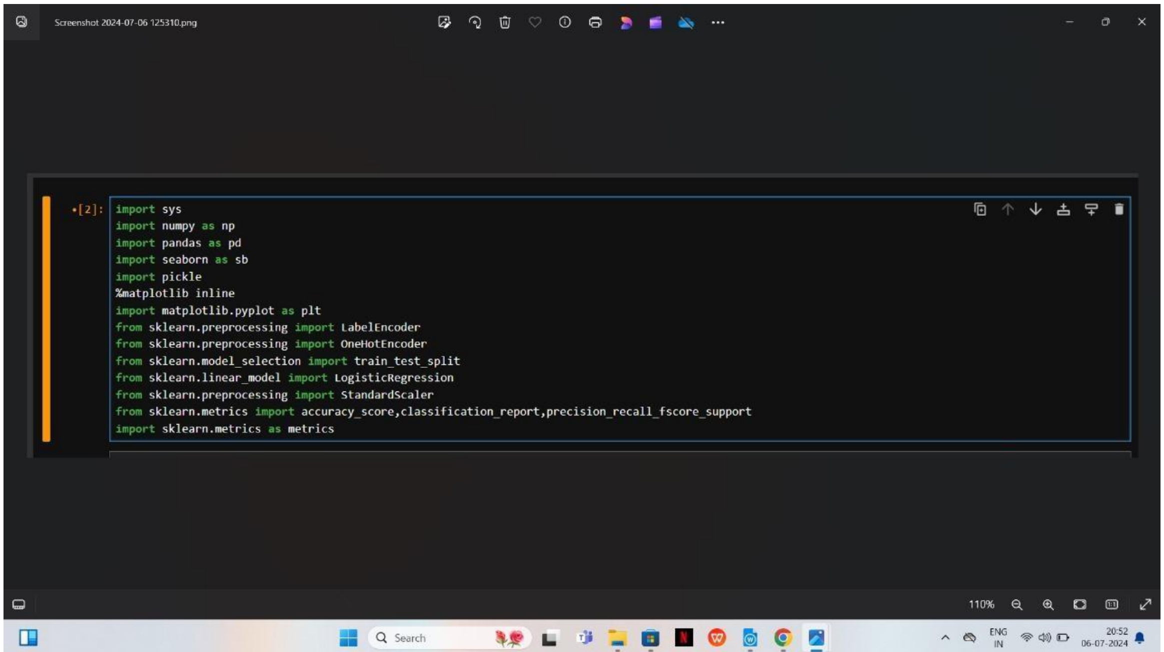
```
from flask import *
import time as time
import pickle
```

```
with open('flight.pkl', 'rb') as file:
   model = pickle.load(file)
   app=Flask(name)


@app.route("/") def home(): return
render_template("home.html")



@app.route("/predict")
def predict():
   return render_template("predict.html")


@app.route("/team") def team():
return render_template("team.html")


@app.route("/result",methods=["POST"])
def result():
   number=int(request.form["fltnum"])
   month=int(request.form["month"])
   dayofmonth=int(request.form["dayofmonth"]
   ) dayofweek=int(request.form["dayofweek"])
   origin=request.form["origin"] if(origin=="atl"):
   origin=0 elif(origin=="dtw"):
      origin=1
   elif(origin=="jfk"):
   origin=2
   elif(origin=="msp"):
```

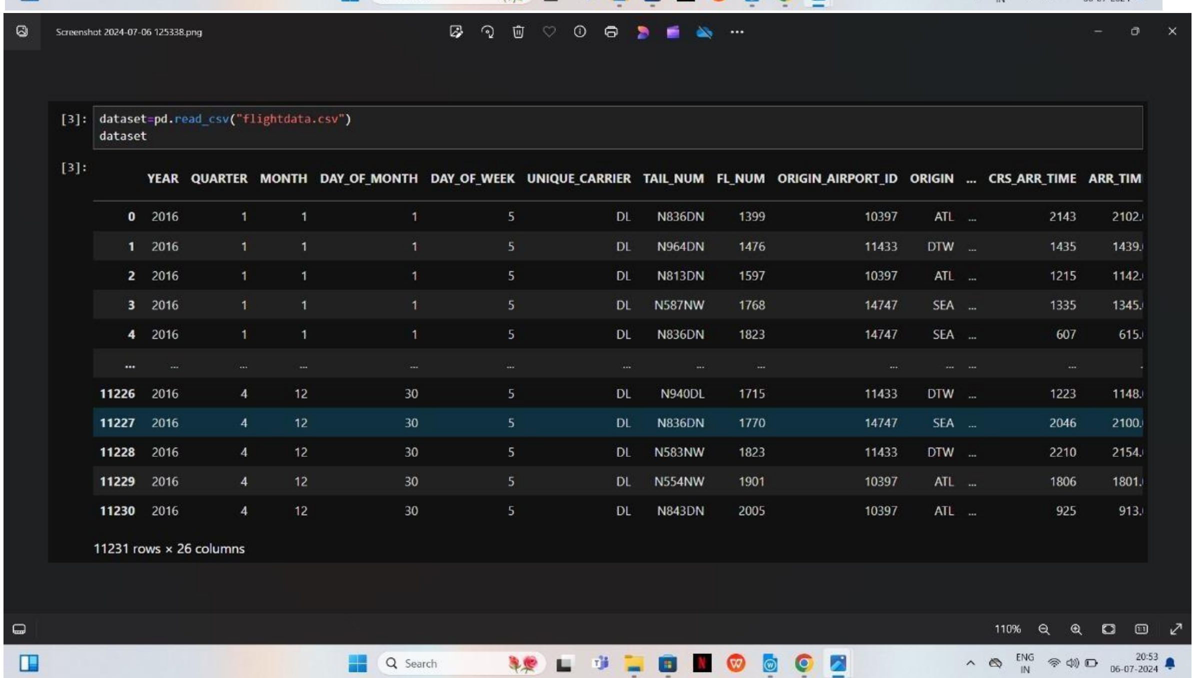
```
    origin=3
    elif(origin=="sea"):
      origin=4
    destination=request.form["destination"]
    if(destination=="atl"): destination=0
    elif(destination=="dtw"): destination=1
    elif(destination=="jfk"): destination=2
    elif(destination=="msp"): destination=3
    elif(destination=="sea"):
       destination=4 dept=int(request.form["dept"])
    arrdept=int(request.form["arrtime"]) actdept=int(request.form["acttime"])
    dept15=int(dept)-int(actdept) print("working")
    total=[[number,month,dayofmonth,dayofweek,origin,destination,dept15,arrdept]]
    y_pred=model.predict(total)

    if(y_pred==[0.]):
       ans="THE FLIGHT WILL BE ON TIME"
       show=0
    else:
       ans="THE FLIGHT WILL BE DELAYED"
       show=1
    print(ans) return
render_template("result.html",ans=ans,show=show) if
name=="main": app.run()
```

CODE SNIPPETS

MODEL BUILDING



```
[2]: import sys
import numpy as np
import pandas as pd
import seaborn as sb
import pickle
%matplotlib inline
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score,classification_report,precision_recall_fscore_support
import sklearn.metrics as metrics
```



```
[3]: dataset=pd.read_csv("flightdata.csv")
dataset
```

[3]:

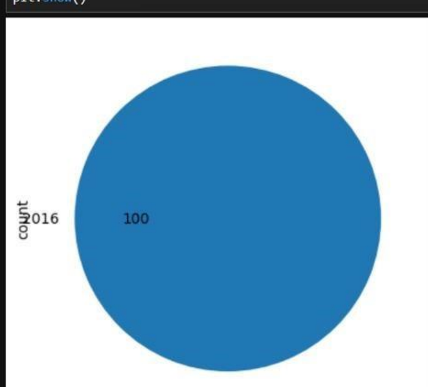
	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	UNIQUE_CARRIER	TAIL_NUM	FL_NUM	ORIGIN_AIRPORT_ID	ORIGIN	...	CRS_ARR_TIME	ARR_TIM
0	2016	1	1	1	5	DL	N836DN	1399	10397	ATL	...	2143	2102.
1	2016	1	1	1	5	DL	N964DN	1476	11433	DTW	...	1435	1439.
2	2016	1	1	1	5	DL	N813DN	1597	10397	ATL	...	1215	1142.
3	2016	1	1	1	5	DL	N587NW	1768	14747	SEA	...	1335	1345.
4	2016	1	1	1	5	DL	N836DN	1823	14747	SEA	...	607	615.
...	...	...	...	...	...	...	...	...	...	...	...	...	.
11226	2016	4	12	30	5	DL	N940DL	1715	11433	DTW	...	1223	1148.
11227	2016	4	12	30	5	DL	N836DN	1770	14747	SEA	...	2046	2100.
11228	2016	4	12	30	5	DL	N583NW	1823	11433	DTW	...	2210	2154.
11229	2016	4	12	30	5	DL	N554NW	1901	10397	ATL	...	1806	1801.
11230	2016	4	12	30	5	DL	N843DN	2005	10397	ATL	...	925	913.

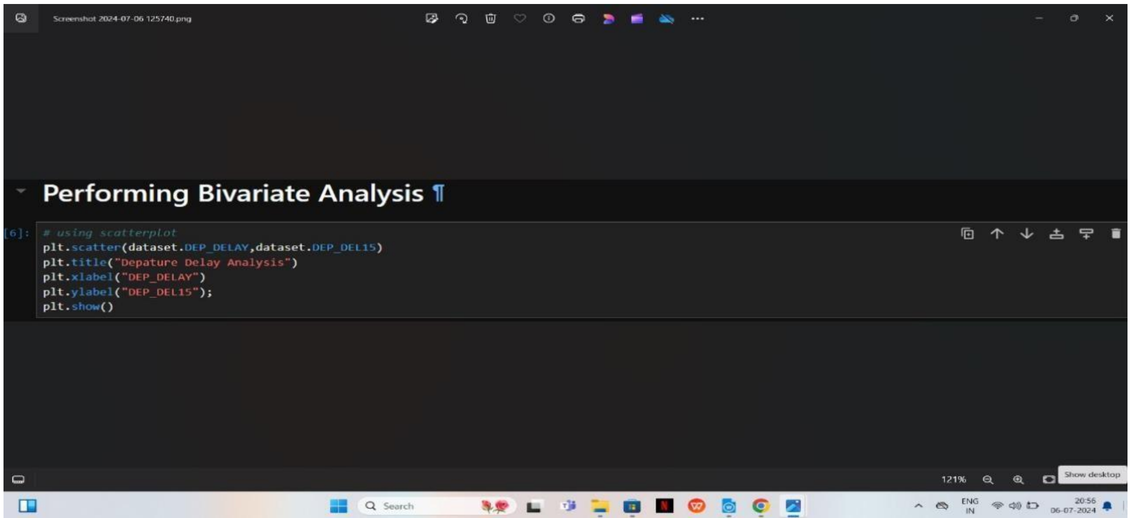
11231 rows × 26 columns

```
[4]: dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11231 entries, 0 to 11230
Data columns (total 26 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   YEAR                11231 non-null  int64
 1   QUARTER             11231 non-null  int64
 2   MONTH               11231 non-null  int64
 3   DAY_OF_MONTH        11231 non-null  int64
 4   DAY_OF_WEEK         11231 non-null  int64
 5   UNIQUE_CARRIER      11231 non-null  object
 6   TAIL_NUM            11231 non-null  object
 7   FL_NUM              11231 non-null  int64
 8   ORIGIN_AIRPORT_ID   11231 non-null  int64
 9   ORIGIN              11231 non-null  object
 10  DEST_AIRPORT_ID     11231 non-null  int64
 11  DEST                11231 non-null  object
 12  CRS_DEP_TIME        11231 non-null  int64
 13  DEP_TIME            11124 non-null  float64
 14  DEP_DELAY           11124 non-null  float64
 15  DEP_DEL15           11124 non-null  float64
 16  CRS_ARR_TIME        11231 non-null  int64
 17  ARR_TIME            11116 non-null  float64
 18  ARR_DELAY           11043 non-null  float64
 19  ARR_DEL15           11043 non-null  float64
 20  CANCELLED           11231 non-null  float64
 21  DIVERTED            11231 non-null  float64
 22  CRS_ELAPSED_TIME    11231 non-null  float64
```

Performing Univariate Analysis

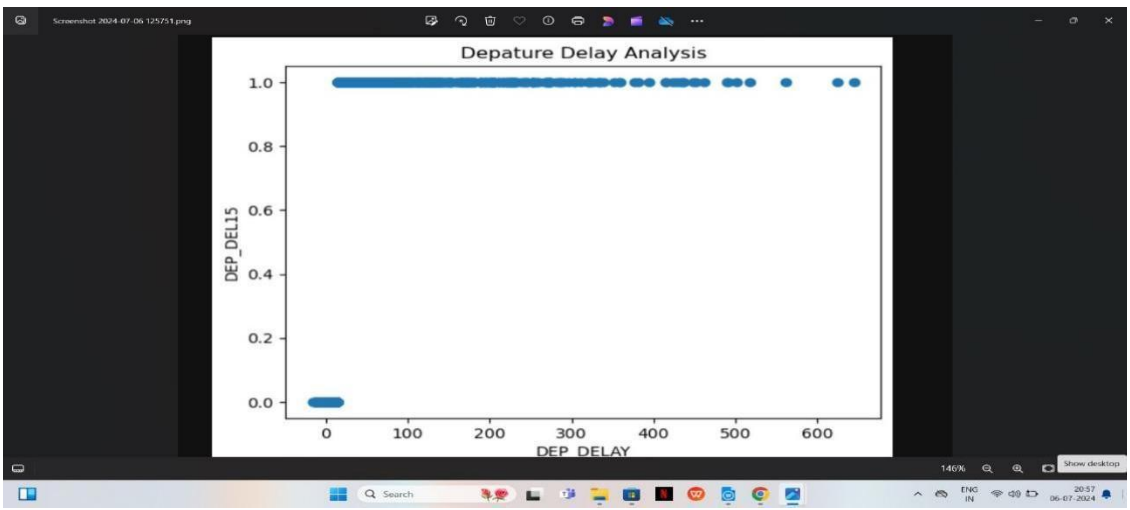
```
[5]: #using pie chart
dataset["YEAR"].value_counts().plot(kind="pie",autopct="%.0f")
plt.show()
```

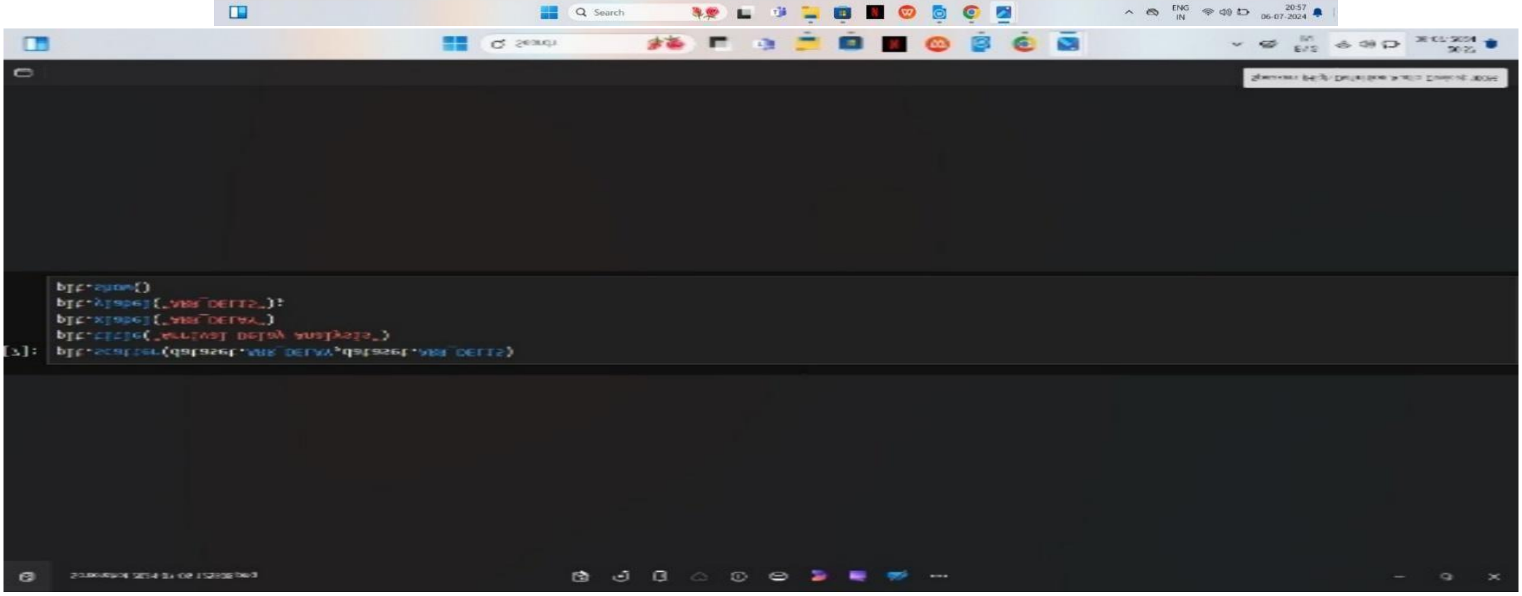




Performing Bivariate Analysis ¶

```
[6]: # using scatterplot
plt.scatter(dataset.DEP_DELAY,dataset.DEP_DEL15)
plt.title("Depature Delay Analysis")
plt.xlabel("DEP_DELAY")
plt.ylabel("DEP_DEL15");
plt.show()
```





Dropping Unnecessary Columns

[95]:
```
dataset=dataset[["FL_NUM","MONTH","DAY_OF_MONTH","DAY_OF_WEEK","ORIGIN","DEST","DEP_DEL15","CRS_ARR_TIME","ARR_DEL15"]]
dataset.head()
```

[95]:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	DEP_DEL15	CRS_ARR_TIME	ARR_DEL15
0	1399	1	1	5	ATL	SEA	0.0	2143	0.0
1	1476	1	1	5	DTW	MSP	0.0	1435	0.0
2	1597	1	1	5	ATL	SEA	0.0	1215	0.0
3	1768	1	1	5	SEA	MSP	0.0	1335	0.0
4	1823	1	1	5	SEA	DTW	0.0	607	0.0

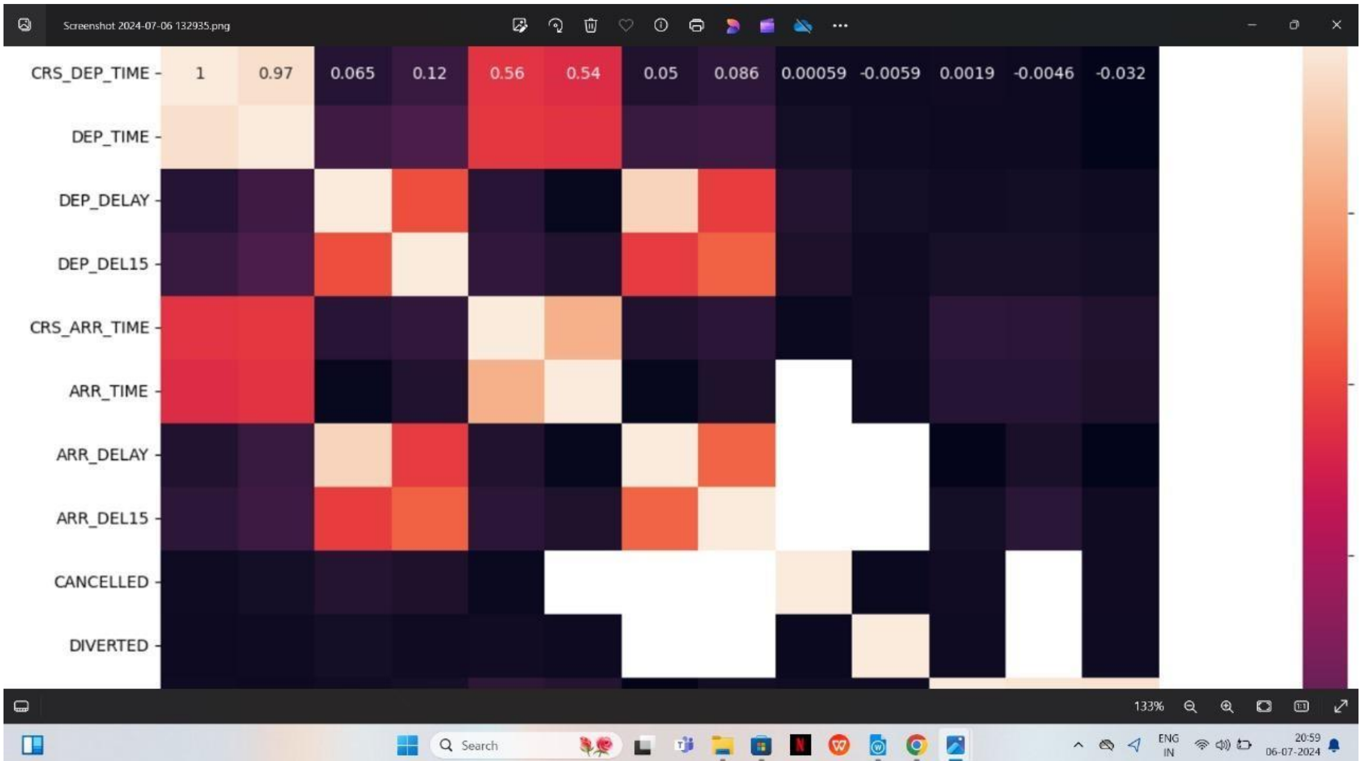
Exploratory Data Analysis(EDA)

94]:
```
dataset.describe()
```

94]:

	YEAR	QUARTER	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	FL_NUM	ORIGIN_AIRPORT_ID	DEST_AIRPORT_ID	CRS_DEP_TIME	DEP_TIME	...
count	11231.0	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11231.000000	11124.000000	...
mean	2016.0	2.544475	6.628973	15.790758	3.960199	1334.325617	12334.516695	12302.274508	1320.798326	1327.189410	...
std	0.0	1.090701	3.354678	8.782056	1.995257	811.875227	1595.026510	1601.988550	490.737845	500.306462	...
min	2016.0	1.000000	1.000000	1.000000	1.000000	7.000000	10397.000000	10397.000000	10.000000	1.000000	...
25%	2016.0	2.000000	4.000000	8.000000	2.000000	624.000000	10397.000000	10397.000000	905.000000	905.000000	...
50%	2016.0	3.000000	7.000000	16.000000	4.000000	1267.000000	12478.000000	12478.000000	1320.000000	1324.000000	...
75%	2016.0	3.000000	9.000000	23.000000	6.000000	2032.000000	13487.000000	13487.000000	1735.000000	1739.000000	...
max	2016.0	4.000000	12.000000	31.000000	7.000000	2853.000000	14747.000000	14747.000000	2359.000000	2400.000000	...

8 rows × 22 columns



Replacing Null values

```
[99]: dataset=dataset.fillna({'DEP_DEL15':dataset['DEP_DEL15'].mode()[0],
                              'ARR_DEL15':dataset["ARR_DEL15"].mode()[0]})
```

```
[100]: dataset.isnull().sum()
```

```
[100]: FL_NUM          0
       MONTH           0
       DAY_OF_MONTH    0
       DAY_OF_WEEK     0
       ORIGIN          0
       DEST            0
       DEP_DEL15       0
       CRS_ARR_TIME    0
       ARR_DEL15       0
       dtype: int64
```

Screenshot 2024-07-06 133304.png

Building The Machine Learning Model

Logistic Regression

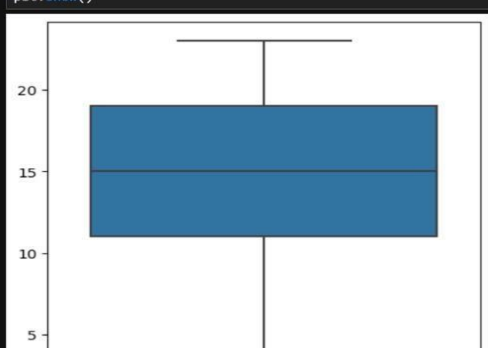
110]:
```
log_reg=LogisticRegression(max_iter=800)
log_reg.fit(X_train,Y_train.ravel())
```

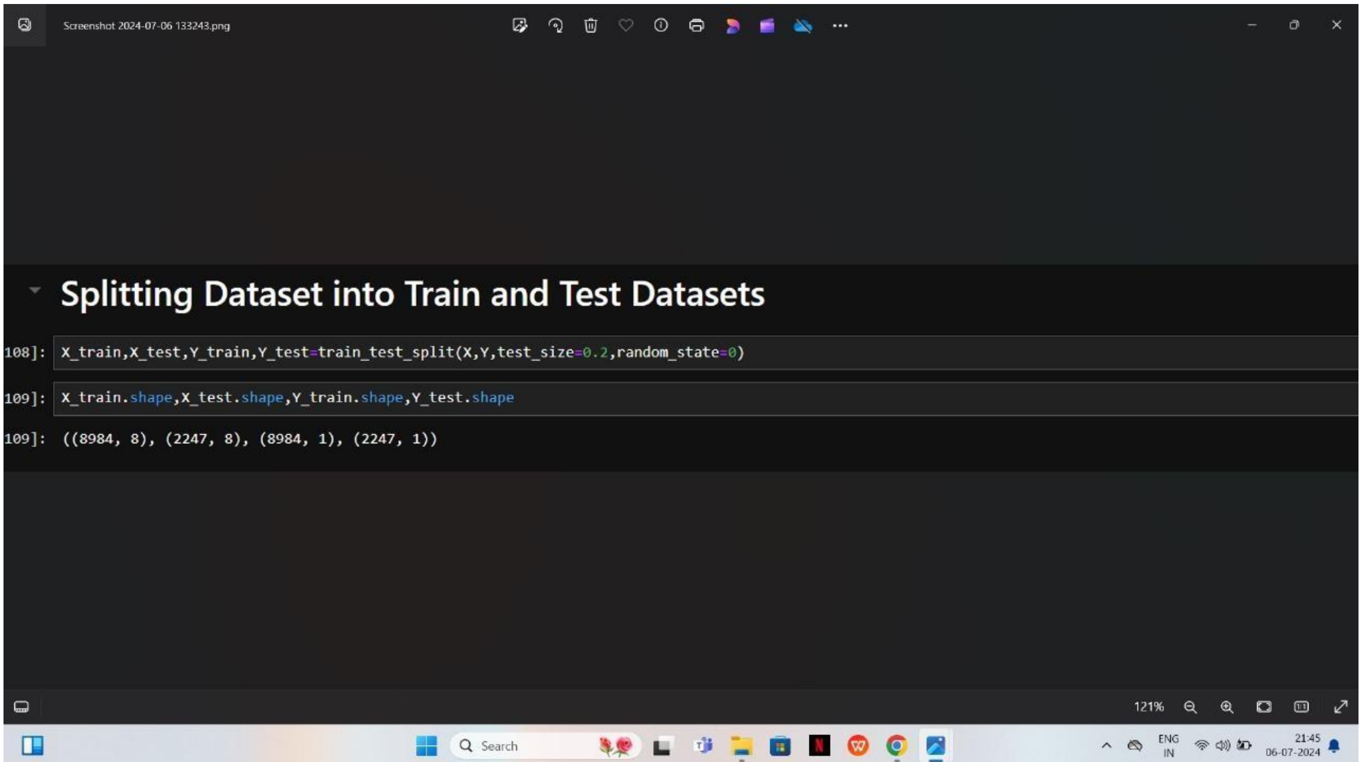
110]:
```
LogisticRegression
LogisticRegression(max_iter=800)
```

Screenshot 2024-07-06 133124.png

Handling Outliers

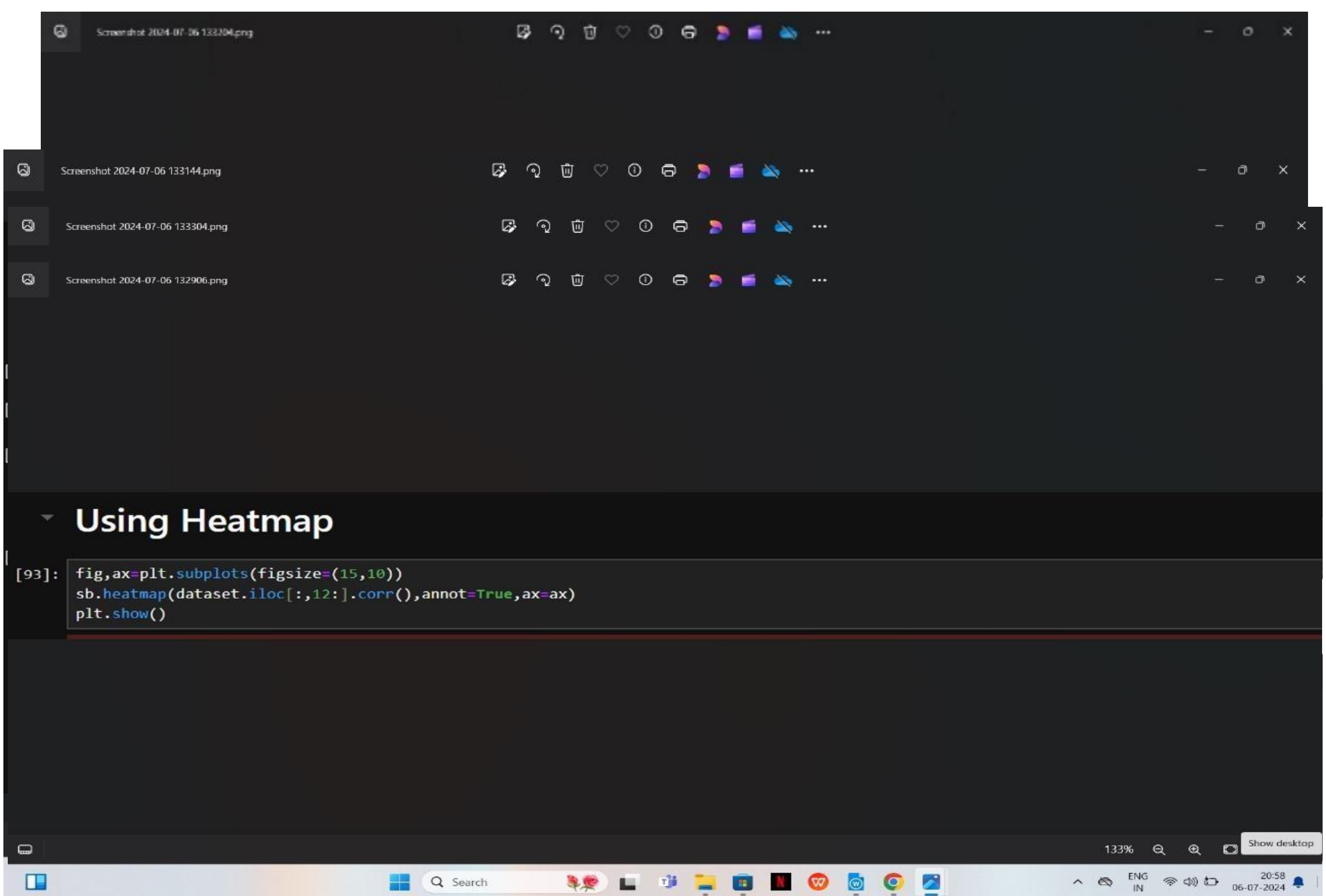
[102]:
```
fig, ax=plt.subplots(figsize=(5,6))
sb.boxplot(data=dataset["CRS_ARR_TIME"])
plt.show()
```





Splitting Dataset into Train and Test Datasets

```
[108]: X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=0)

[109]: X_train.shape,X_test.shape,Y_train.shape,Y_test.shape

[109]: ((8984, 8), (2247, 8), (8984, 1), (2247, 1))
```



Using Heatmap

```
[93]: fig,ax=plt.subplots(figsize=(15,10))
      sb.heatmap(dataset.iloc[:,12:].corr(),annot=True,ax=ax)
      plt.show()
```

Screenshot 2024-07-06 133425.png

Checking for Overfitting and Underfitting

```
[115]: log_train_acc=accuracy_score(Y_train,Y_pred_log_train)
       log_test_acc=accuracy_score(Y_test,Y_pred_log_test)
       print("Training Accuracy=",log_train_acc)
       print("Testing Accuracy=",log_test_acc)

       Training Accuracy= 0.9182991985752449
       Testing Accuracy= 0.9287939474855362
```

There is no big variation between training and testing accuracy. Therefore the Logistic Regression model is not overfited and underfited

Screenshot 2024-07-06 133446.png

Saving The Model

```
[117]: pickle.dump(log_reg,open("flight.pkl","wb"))
```

Screenshot 2024-07-06 133322.png

Testing The Model

```
[111]: Y_pred_log_train=log_reg.predict(X_train)
       Y_pred_log_test=log_reg.predict(X_test)

[112]: pd.DataFrame(Y_pred_log_train).value_counts()

[112]: 0.0    7705
       1.0    1279
       Name: count, dtype: int64
```

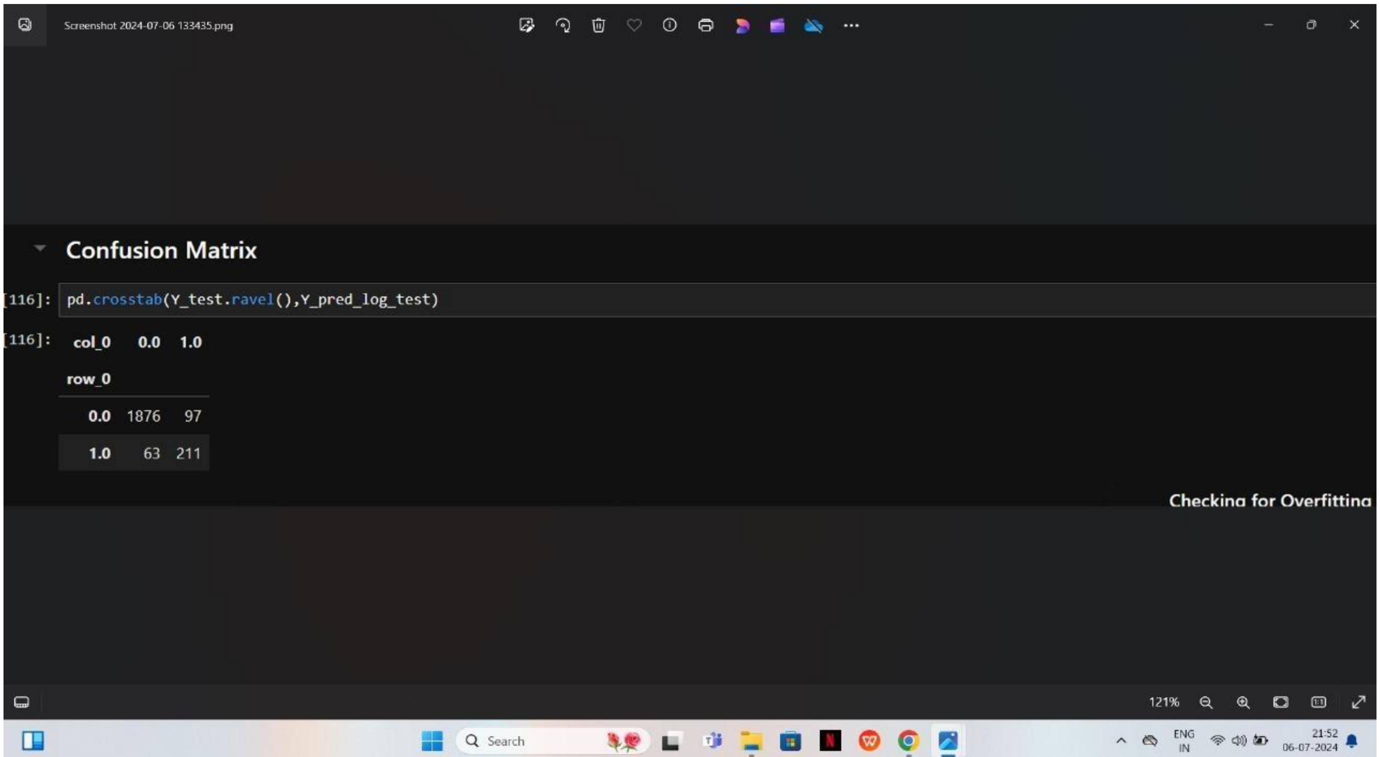
Screenshot 2024-07-06 133413.png

Evaluating The Model Using Metrics

Classification Report

```
13]: print(classification_report(Y_test,Y_pred_log_test))

              precision    recall  f1-score   support

         0.0       0.97      0.95      0.96      1973
         1.0       0.69      0.77      0.73       274

    accuracy                           0.93      2247
   macro avg       0.83      0.86      0.84      2247
weighted avg       0.93      0.93      0.93      2247
```

```
14]: acc_log=accuracy_score(Y_test,Y_pred_log_test)
     prec_log,rec_log,f1_log,sup_log=precision_recall_fscore_support(Y_test,Y_pred_log_test)
     print("Accuracy score=",acc_log)
     print("Precision=",prec_log[0])
     print("Recall=",rec_log[0])
     print("F1 score=",f1_log[0])

     Accuracy score= 0.9287939474855362
     Precision= 0.9675090252707581
     Recall= 0.9508362899138368
     F1 score= 0.9591002044989776
```


Screenshot 2024-07-06 133435.png
Confusion Matrix
[116]: pd.crosstab(Y_test.ravel(),Y_pred_log_test)
[116]: col_0 0.0 1.0
row_0
0.0 1876 97
1.0 63 211
Checking for Overfitting
121%
Search
ENG
IN
21:52
06-07-2024