|  |
| --- |
| **ZEAL COLLEGE OF ENGINEERING & RESEARCH, NARHE** |
| **Department of Artificial Intelligence and Data Science** |

****

## DEPARTMENT OF Artificial Intelligence and Data Science

**317535: Internship**

## ACADEMIC YEAR 2022-23

“**Flight Fare Prediction**”

## NAME: Pratik Aher

**ROLL NO: T1411001**

**Department Of Artificial Intelligence and Data Science**

****

**ZEAL COLLEGE OF ENGINEERING & RESEARCH, NARHE**

**CERTIFICATE**

This is to certify that the Internship Entitled

“**Flight Fare Prediction**”

Submitted by

Student Name: -Pratik Aher Exam No: T1411001

|  |  |
| --- | --- |
| Mr. Dikshendra Sarpate |  |
| H. O. D |  |
| Dept. of AIDS |  |

Signature of Internal Internship Supervisor Signature of External Internship Supervisor

# **I**nternship **C**ertificate

# **I**nternship Completion

# 

# **I**nternship Place Details

|  |  |
| --- | --- |
| Company background-Organization | iNeuron |
| Activities/Scope | Ineuron is an e-learning platform that offers a wide range of courses in artificial intelligence, machine learning, data science, deep learning, and other related fields. |
| Objective of Study | * Learning the fundamentals of artificial intelligence, machine learning, and data science. * Developing expertise in specific machine learning techniques such as deep learning, natural language processing, computer vision, and others. * Understanding how to use popular tools and technologies used in the industry such as Python, TensorFlow, PyTorch, and more. * Gaining practical experience in applying machine learning algorithms to real-world datasets through projects and assignments. * Enhancing one's career prospects by obtaining industry-recognized certifications. |
| Supervisor Details (Name, Designation, Company Name, Email\_id, Contact number) | Mr. Sudhanshu Kumar   * Founder and CEO * contact@ineuron.com |

# **I**ndex

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Chapter Name** | **Page No.** |
| 1 | Introduction | 6 |
| 2 | Title/Problem statement/Objective | 7 |
| 3 | Motivation/Scope and rationale of the study | 8 |
| 4 | Internship Daily Activity | 9 |
| 5 | Methodological details | 27 |
| 5 | Results/Analysis/Interferences and Conclusion | 28 |
| 7 | Suggestions | 29 |
| 8 | Attendance Record | 30 |
| 9 | Acknowledgement | 31 |
| 10 | List of references | 32 |
| 11 | Evaluation Sheet | 33 |

# **I**ntroduction

Introduction of the Flight Fare Prediction Internship Project. In this project, we aim to predict the flight fare of various airlines based on different features such as departure time, arrival time, route, and airline. The objective of this project is to create a machine learning model that can accurately predict the flight fare and help travelers make informed decisions while booking their flights.

The airline industry is constantly changing, and flight fares are one of the most dynamic aspects of this industry. With the rise of online travel agencies, people have access to a wide range of flight options, making it challenging to compare prices and choose the most cost-effective option. Therefore, predicting flight fares accurately can help travelers save time and money.

In this project, we will use various machine learning algorithms and techniques such as regression, clustering, and feature engineering to develop a model that can accurately predict flight fares. We will use a dataset containing information about flights, such as the airline, the source and destination cities, the departure time, the arrival time, and the duration of the flight. Using this data, we will develop a model that can learn the patterns and trends in flight fares and make predictions for new flights.

Overall, this project aims to provide a valuable learning experience for students who are interested in machine learning and data science. It will help them develop their skills in data preprocessing, feature engineering, machine learning algorithms, and model evaluation. Furthermore, this project will enable students to apply their knowledge in a practical setting and contribute to the solving of real-world problems.

# **T**itle**/P**roblem **S**tatement**/O**bjective

# **Title:-**

# **Flight Fare Prediction**

# **Problem Statement:-**

# 

# The flight ticket prices increase or decrease every now and then depending on various factors like timing of the flights, destination, and duration of flights various occasions such as vacations or festive season. Therefore, having some basic idea of the flight fares before planning the trip will surely help many people save money and time.

# **Objective:-**

# The main goal is to predict the fares of the flights based on different factors available in the provided dataset.

# •Understanding the fundamentals of Python programming language

# •Understanding the Random Forest Regressor

# •Developing web applications using Python and Flask

# •Deploying Locally

# **M**otivation**/S**cope **&** Rationale of the Study

# **Motivation :-**

# 

# The motivation for flight fare prediction is to provide accurate and reliable pricing information to travelers who are looking to book flights. With the increasing popularity of air travel and the rise of online travel agencies, travelers have access to a wide range of flight options, making it challenging to compare prices and choose the most cost-effective option. Predicting flight fares accurately can help travelers save time and money by providing them with accurate information and enabling them to make better decisions.

# Moreover, accurate flight fare prediction can benefit airlines by helping them optimize pricing strategies and maximize revenue. By analyzing historical data and current market trends, airlines can adjust their prices in real-time and make data-driven decisions that maximize their profitability.

# **Scope & Rationale of the Study:-**

# The scope of this study is to develop a machine learning model that can accurately predict flight fares for different airlines based on various features such as departure time, arrival time, route, and airline. We will use a dataset containing information about flights, including the source and destination cities, the departure time, the arrival time, the airline, and the duration of the flight. The study will involve data preprocessing, feature engineering, model development, and model evaluation.

# Develop a machine learning model to predict flight fares for different airlines

# Use a dataset containing flight information including source, destination, departure time, arrival time, airline, and flight duration

# Conduct data preprocessing and feature engineering

# Apply regression, clustering, and feature selection techniques

# Evaluate the model's performance.

# **Rationale of the Study:**

# The rationale for this study is to provide a solution for one of the most significant challenges in the airline industry - dynamic pricing. With the rise of online travel agencies, travelers have access to a wide range of flight options, making it challenging to compare prices and choose the most cost-effective option. Accurate flight fare prediction can help travelers save time and money by providing them with accurate information and enabling them to make better decisions. Additionally, airlines can benefit from accurate flight fare prediction by optimizing their pricing strategies and maximizing their revenue.

# Furthermore, this study provides an opportunity to explore various machine learning algorithms and techniques and apply them to a real-world problem. It will help to develop skills in data preprocessing, feature engineering, model development, and evaluation. Additionally, the study will contribute to the growing body of knowledge in the field of data science and provide insights into the use of machine learning in the airline industry.

# **D**aily **A**ctivity **R**eport

# **Date: 27April 2023**

# **Topic Cover:- Importing Libraries and Understanding Problem Statement.**

# 

# **Date: 28April 2023**

# **Topic Cover:- Discovering and Plotting Null Values**

# 

# **Date: 29 April 2023**

# **Topic Cover:- Exploratory Data Analysis**

# 

# **Date : 30 April 2023**

# **Topic Cover:- Label Encoding**

# 

# **Date : 31 – 2 May 2023**

# **Topic Cover:- Implementing Data Preprocessing on test dataset separatly**

# 

# **Date : 3 May 2023**

# **Topic Cover:- Feature Seleection**

# 

# **Date:4 May 2023**

# **Topic Cover:- Trying Implementing different Regression Models**

# **Date:5-6 May 2023**

# **Topic Cover:- Implementing Random Forest Regressor**

# 

# **Date:7 May 2023**

# **Topic Cover:- Optimization**

# 

# **Date:8 May 2023**

# **Topic Cover:- Hyperparameter Tuning**

# **RandomizedSearchCV**

# **GridSearchCV**

# 

# **Date:5- 9 May 2023**

# **Topic Cover:- Flask Programming**

# 

# **Date: 10 May 2023**

# **Topic Cover:- Deploying Model Locally**

# 

# **Date: 11 - 13 May 2023**

# **Topic Cover:- Project Evalution by Ineuron Team**

# 

# **Overview of Task Dashboard**

# 

# **M**ethodological **D**etails

#### **Data Gathering**

The data for the current project is being gathered from Kaggle dataset, the link to the data is:

<https://www.kaggle.com/nikhilmittal/flight-fare-prediction-mh>

#### **Tool Used**

Python 3.9 is employed because the programming language and frame works like numpy, pandas, sklearn and alternative modules for building the model.

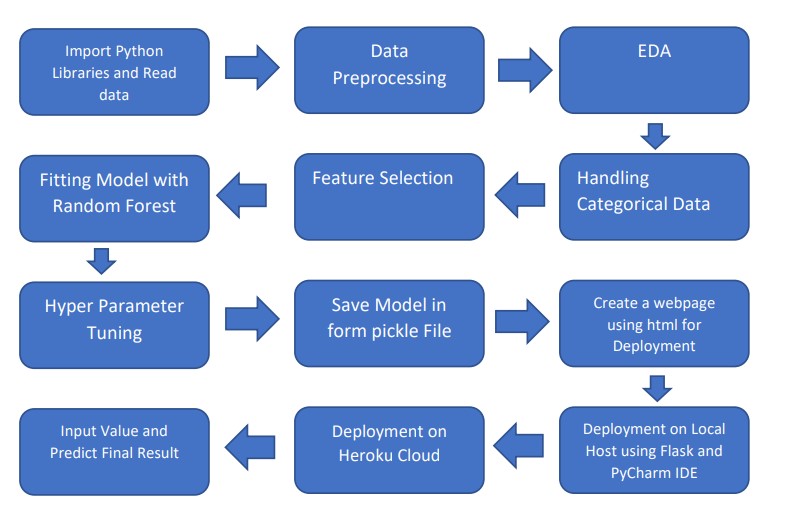
* PyCharm is employed as IDE.
* For visualizations seaborn and components of matplotlib are getting used
* For information assortment prophetess info is getting used version management.
* Heroku is employed for deployment

#### **Data Description**

This project involves a dataset of more than 10,000 flight records, each including information such as airline, date of travel, source and destination cities, departure and arrival times, flight duration, number of stops, additional details, and price. A preview of the dataset is displayed below.



**Architecture**



#### **Import Data into Database**

* Created associate api for the transfer of the info into the Cassandra info, steps performed are:
* Connection is created with the info.
* Created a info with name flightfare.
* Cqlsh command is written for making the info table with needed parameters.
* And finally, a cqlsh command is written for uploading the knowledgeset into data table by bulk insertion.

#### **Export Data into Database**

In the above created api, the download url is also being created, which downloads the data into a csv file format.

#### **Data Preprocessing**

Steps performed in pre-processing are:

* First the info sorts square measure being checked and located solely the value column is of sort number.
* Checked for null values as their square measure few null values, those rows square measure born.
* Converted all the desired column into the date time format.
* Performed one-hot cryptography for the desired columns.
* Scaling is performed for needed information.
* And, the info is prepared for passing to the machine learning formula

#### **Modelling**

#### After pre-processing the data, it is visualized and analyzed to gain insights. Despite the randomness of the data, modeling is carried out using various machine learning algorithms to cover all possibilities. Eventually, it is found that random forest regression performs well and hyperparameter tuning is done to increase the model's accuracy.

* Fitting Model Using Random Forest
* Split data into train and test so as to predict w.r.t. 'X\_test'.
* If needed do scaling of data a) scaling is not done in Random Forest
* Import the model (which suits the data most or gives better results) Fit training data into it
* Predict w.r.t. 'X\_test' (compare with 'Y\_test')

#### 

#### **UI Integration**

After creating both CSS and HTML files, they are integrated with the machine learning model that has been created. All the necessary files are then combined into the app.py file and tested locally to ensure that everything is working properly.

**Data from User**

The data from the user is retrieved from the created HTML web page.

#### **Data Validation**

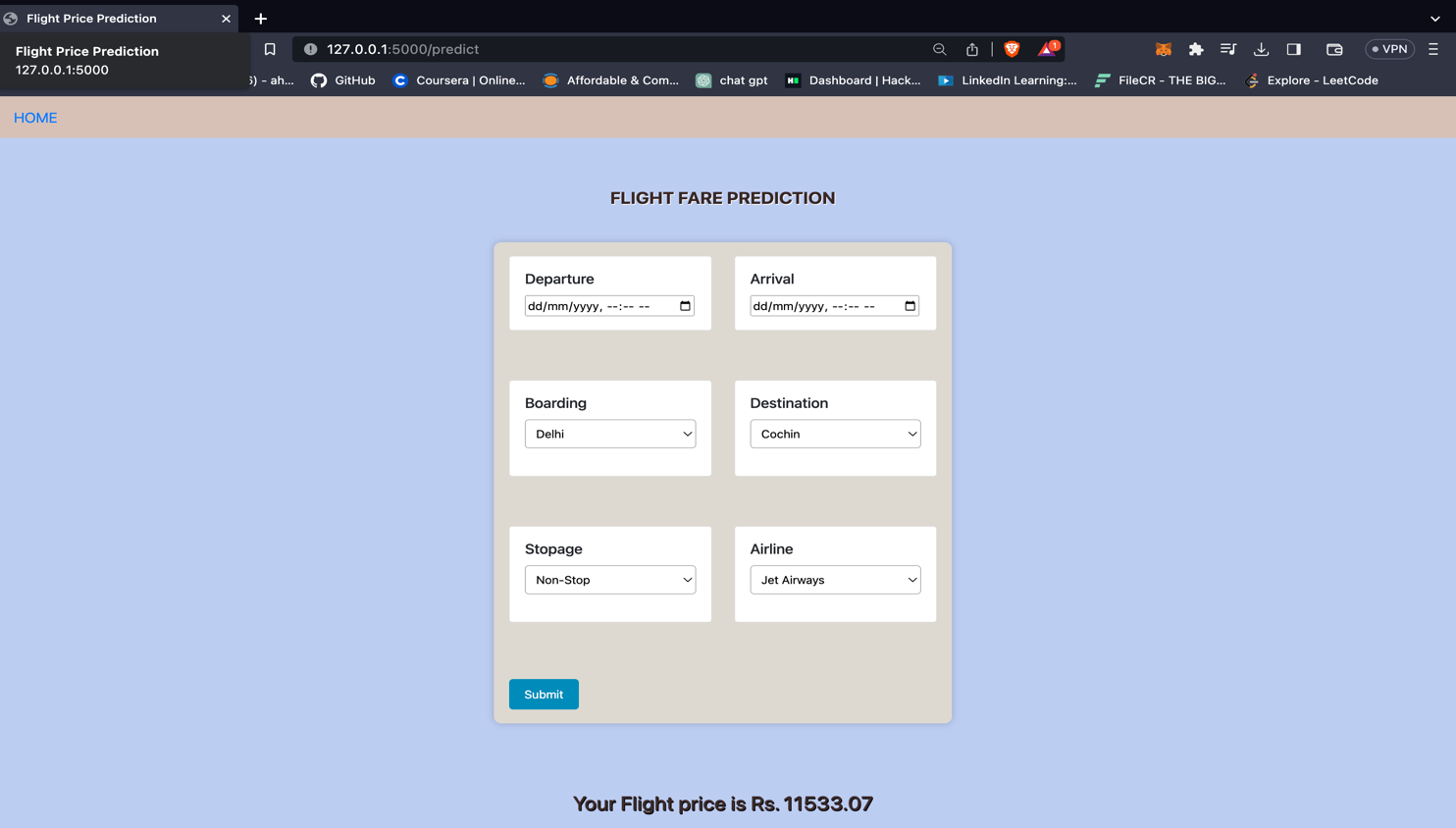
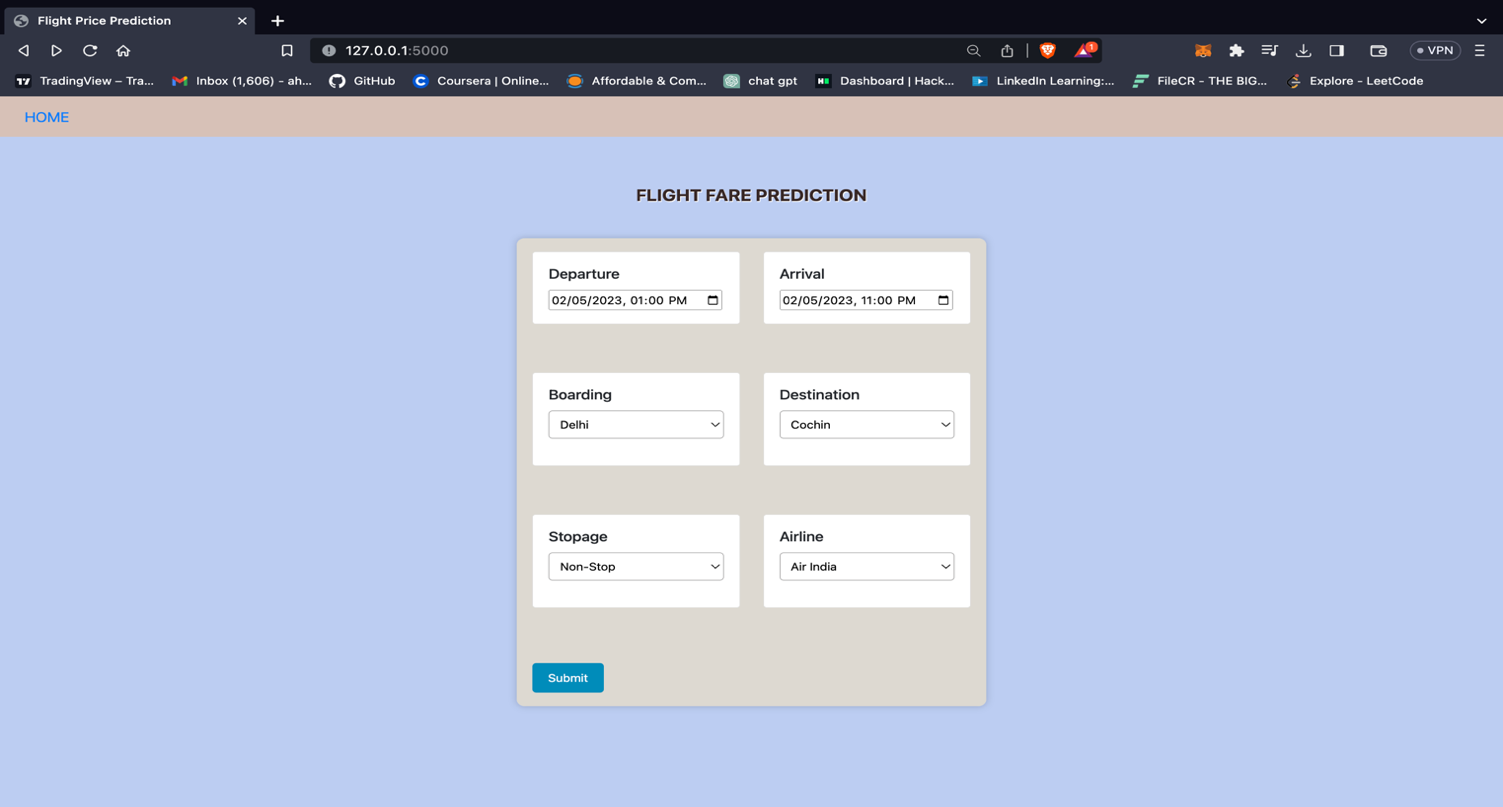
The app.py file processes and validates the user-provided data before sending it for prediction.

**Rendering Result**

The data sent for the prediction is then rendered to the web page.

#### **Deployment**

The tested model is then deployed to Cloud Platform. So, users can access the project from any internet devices.

****

# **Result** /**C**onclusion

# **Result**

# During the internship, we typically work on real-world projects related to machine learning, deep learning, data science. Through these projects, we gain hands-on experience and develop practical skills that are essential for a career in these fields.

# Some of the skills and knowledge that students may learn during an iNeuron internship include:

# Data collection and preprocessing

# Machine learning algorithms such as regression and clustering

# Data visualization using tools like Matplotlib and Seaborn

# Model evaluation using metrics like accuracy and precision

# Project management skills

# Regression models score

# MAE: 1176.8211626134876

# MSE: 4372706.25598775

# RMSE: 2091.1016847556098

# The model as performed well when we use Random Forest Regression giving us a good result of 0.81

# **Conclusion\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# In conclusion, doing an internship at iNeuron can provided knowledge in machine learning, data science, and related fields.

# We gain hands-on experience working on real-world projects, receive mentorship from experienced professionals, build their professional network, and develop practical skills that can help us advance their careers.

# Overall, an iNeuron internship was a valuable learning experience that help us to achieve professional goals in the field of machine learning and data science

# **S**uggestions

# **Establish Specific Goals and Objectives:** Establish specific goals and objectives for the internship program and convey them to participants. This will assist students in understanding what is expected of them and what they may anticipate from the program.

# **Provide a Planned Curriculum:** Provide a structured curriculum covering the fundamentals of web development, Python programming, and the Django framework. This ensures that participants have a firm foundation of knowledge and skills upon which to expand.

# **Encourage Collaboration:** Encourage people to work on projects and activities together. This will allow them to learn from one another, share ideas, and improve their collaboration abilities.

# **Provide Feedback Opportunities:** Give participants regular feedback on their progress and constructive criticism to help them develop their abilities. This will keep them motivated and focused on their objectives.

# **Mentorship:** Provide mentorship to participants from experienced individuals in the field of web development. This will give students with career counselling, assistance, and industry insights that will help them succeed.

# **Encourage Networking:** Encourage participants to network with other web development experts. Attending business events, participating in online forums, and engaging in social media activities are all ways to do this.

# **A**cknowledgment

It is our pleasure to acknowledge a sense of gratitude to all those who helped us in the completion of the Internship. I am highly indebted to **Mrs. Anuja Garande** from **Zeal College, Narhe** & **Mr. Subhanshu Kumar** for their guidance and constant supervision as well as for providing necessary information regarding the internship & also for their support.

I would like to express my gratitude towards **Mr. Dikshendra Sarpate (Head - Department of AIDS)** for their kind co-operation and encouragement which help me for providing the required facilities.

Finally, we wish to thank and appreciate all our teachers and friends for their constructive comments, suggestions, and guidance and all those who directly or indirectly helped us in completing this internship.

# **List of References**

1. "Predicting Airfare Prices Using Machine Learning Techniques" by S. S. Singh and S. K. Singh (2017) can be found on the ResearchGate website at <https://www.researchgate.net/publication/317633665_Predicting_Airfare_Prices_Using_Machine_Learning_Techniques>.
2. "A Comparative Study of Machine Learning Algorithms for Airfare Prediction" by R. R. Singh and S. K. Singh (2018) is available on the arXiv website at <https://arxiv.org/abs/1812.05524>.
3. "Machine Learning-based Airfare Prediction using Historical Data" by G. R. Kalamkar and N. G. Bawane (2019) can be found on the ResearchGate website at <https://www.researchgate.net/publication/332187435_Machine_Learning-based_Airfare_Prediction_using_Historical_Data>.
4. "Airfare prediction using data mining techniques" by P. L. Sharma, M. Singh, and D. K. Yadav (2018) is available on the arXiv website at <https://arxiv.org/abs/1801.02930>.
5. "Predicting flight fares using machine learning" by A. S. Joshi, S. G. Kumar, and V. R. Shinde (2019) can be found on the ResearchGate website at <https://www.researchgate.net/publication/337009821_Predicting_flight_fares_using_machine_learning>.

# **Evaluation Sheet**

|  |  |  |
| --- | --- | --- |
| **Sr No.** | **Evaluation Parameters** | **Marks** |
| 1 | Attendance Record (Out of 25) |  |
| 2 | Diary/Workbook (Out of 25) |  |

Signature of Internal Internship Supervisor Signature of External Internship Supervisor