

INDUSTRIAL TRAINING REPORT

ON

“Flight Price Prediction”

Submitted in partial fulfillment of the requirements
for the award of degree of

Bachelor of technology

In

INFORMATION TECHNOLOGY

Submitted To: -

Shallu Bashambu

Submitted By: - Saransh Kaushik

07614803118



MAHARAJA AGRASEN INSTITUTE OF TECHNOLOGY

DELHI -110086

Affiliated To

GURU GOBIND SINGH INDRAAPRASTHA UNIVERSITY

DECLARATION

I hereby declare that the training report entitled “Flight Price Prediction” is an authentic report of my own work as requirements of 4 weeks training during the period from July 9, 2021 to Aug 9, 2021 for the award of degree of B.tech (Information Technology), Maharaja Agrasen Institute of Technology, GGSIPU.

I further declare that this report is based on the information collected by me and whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them in the text of the report and giving their details in the references.

Submitted by –

Saransh Kaushik
07614803118
7th Semester IT
Maharaja Agrasen Institute of Technology

ACKNOWLEDGEMENT

I, Saransh Kaushik (07614803118) a student of 7th semester IT branch of Maharaja Agrasen Institute of Technology, Rohini would like to convey my gratitude to the Management, Director, HOD (IT) Dr M.L. Sharma and Ms Shallu Bashambu (Assistant Professor) of MAIT for emphasizing on 4-6 weeks Industrial Training Program and giving me the platform to interact with industry professionals.

I would also like to thank Mr. Anuj Garg Founder and Mentor –Code for Cause for giving me the opportunity to learn so much in their prestigious organization.

I extend my warm gratitude and regards to everyone who helped me during my internship for their valuable guidance, support and encouragement.

Saransh Kaushik
University Roll No.07614803118

CERTIFICATE



Certificate Of Excellence

PRESENTED TO

Saransh Kaushik

For Completing The Assignments & Working On
Real-World Projects In The
30 Days Machine Learning Bootcamp

A handwritten signature in black ink.

Anuj Garg

Founder @ Code for Cause

A handwritten signature in black ink.

Kunal Kushwaha

Co-founder @ Code for Cause

09 July - 09 August 2021

ABSTRACT

IT industry is useful in changing past and present. In recent times many places are attracting a lot of travellers. Each year the airline industry is considered as one of the most sophisticated industry in using complex pricing strategies. Nowadays flight prices are on point and predictable that it prizes change frequently customers are seeking to get the lowest price for advocates while airline companies are trying to keep their overall revenue as well as possible using technology.

It is possible to reduce the uncertainty of flight rider prices so here we will take predicting the flight prices using efficient machine learning techniques so that customers can also predict and choose the best prize suited to them to reach their desired destination.

Considering some of the following parameters -distance of travel, number of stops, airline, date of journey, arrival time, departure time source etc. , a model is prepared which can predict the cost journey through the flight of a specific airline with respect to time and date.

CONTENTS

DECLARATION	2
ACKNOWLEDGEMENT	3
CERTIFICATE	4
ABSTRACT	5
CONTENTS	6
LIST OF FIGURES	7
LIST OF TABLES	8

LIST OF FIGURES

1. Organisation Logo	9
2. Different Python Libraries	11
3. Applications of Machine learning	12
4. Anaconda Navigator pane	13
5. Jupyter logo	14
6. Model Flowchart	17

LIST OF TABLES

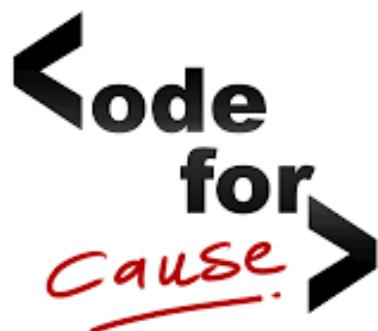
1. Introduction to Organization	9
1.1 Code for cause	
2. Software training work undertaken	
2.1 Machine Learning (Tools learned)	10
3. Project Work	
3.1 Machine Learning Software tools learned	13
3.2 Methodology Used	15
4. Results and Discussion	18
5. Conclusion and future Scope	19
5.1. Conclusion	
5.2. Future Scope	
6. References	20

1 - INTRODUCTION OF THE ORGANISATIONS

1.1 Code for cause:

The organization is an online platform to help the community by providing training, guidance and awareness about the possibilities in the software field to students & professionals. It aims to provide a real world practical learning experience and keep students informed about the latest trends in technology, open-source and opportunities, so that they can keep up with the fast-paced digital world by following a pi-shaped learning pattern.

The organization is aimed to provide learning to students in the best and cheapest way possible as their firm believe is that every student should be able to learn and access new things.



2. SOFTWARE TRAINING WORK UNDERTAKEN

2.1 MACHINE LEARNING

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: ***The ability to learn.*** Machine learning is actively being used today, perhaps in many more places than one would expect.

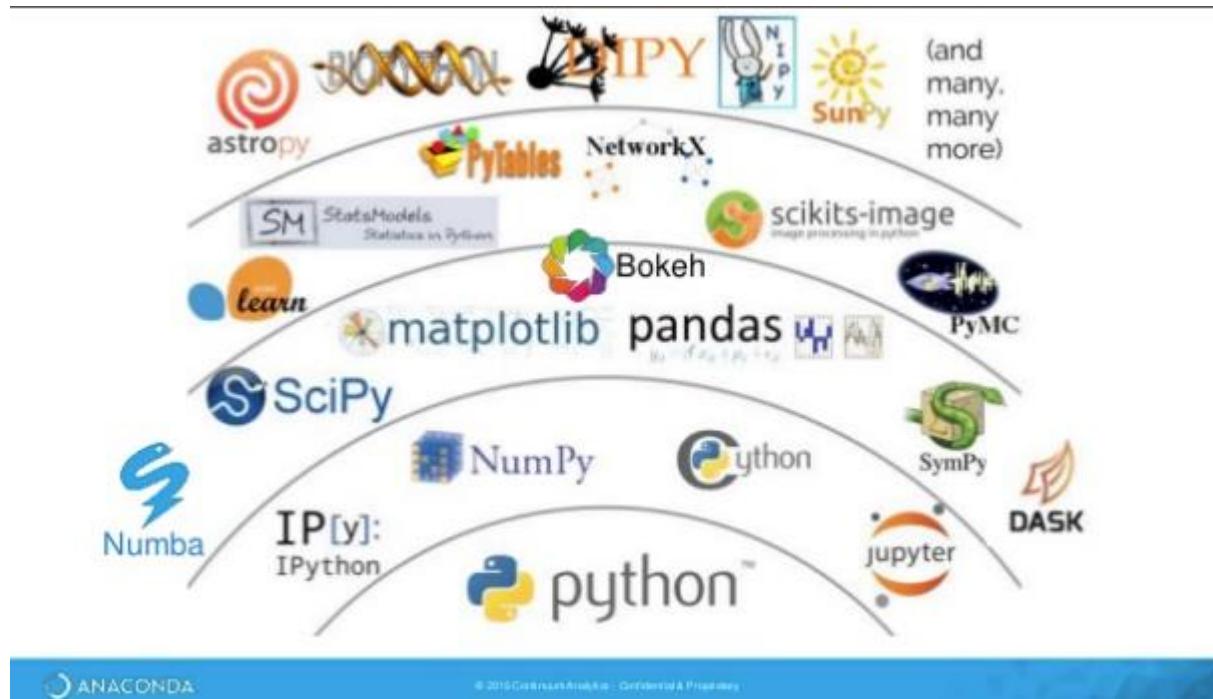
The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that “it gives computers the ability to learn without being explicitly programmed”.

And in 1997, Tom Mitchell gave a “well-posed” mathematical and relational definition that “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”

Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in Machine Learning are:

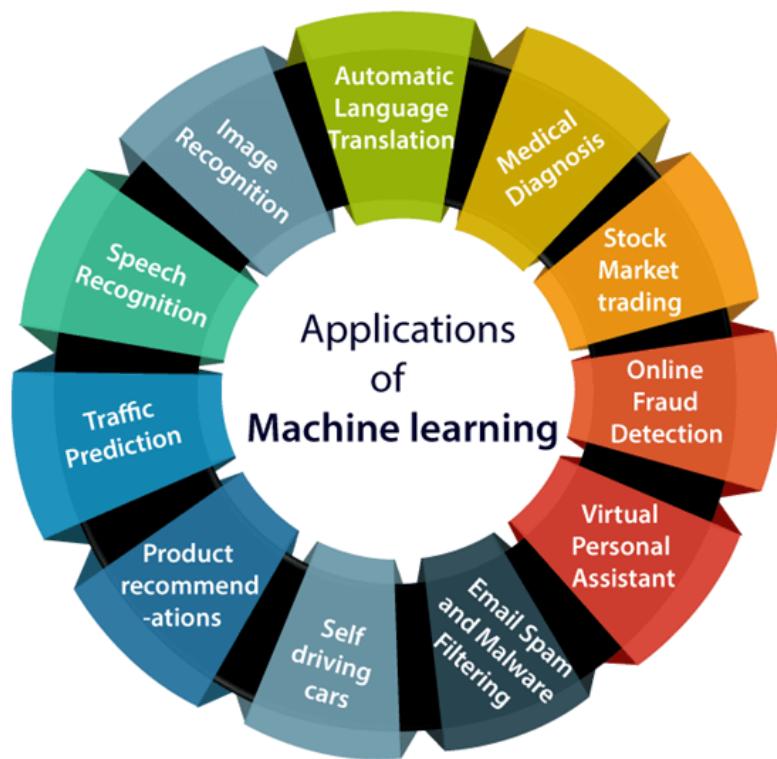
- Numpy
- Scipy

- Scikit-learn
- Theano
- TensorFlow
- Keras
- PyTorch
- Pandas
- Matplotlib



Applications of Machine learning

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:

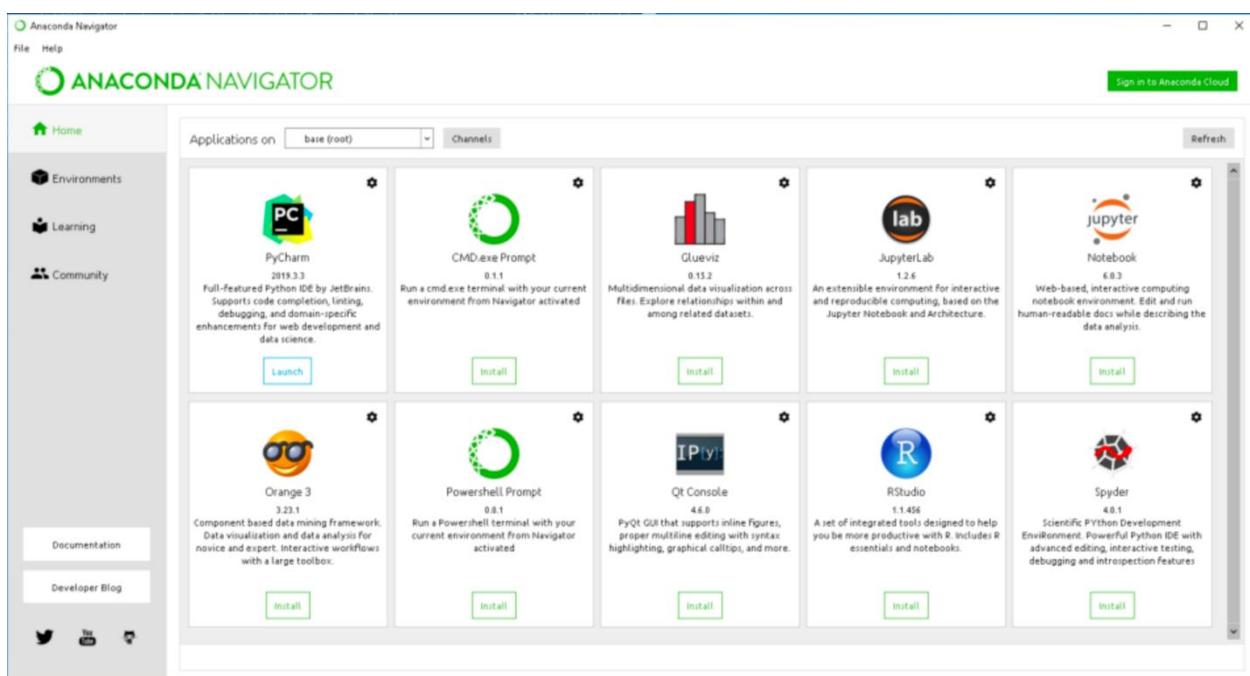


3 PROJECT WORK

3.1 MACHINE LEARNING

SOFTWARE TOOLS LEARNED

- **Anaconda:**



Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS.

- **Jupyter:**



Project Jupyter is a nonprofit organization created to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages". Spun off from IPython in 2014 by Fernando Pérez, Project Jupyter supports execution environments in several dozen languages.

3.2 Methodology Used

Machine Learning Models used in this project-

1.Linear Regression

Simple linear regression is a statistical method that allows us to summarise and study relationships between two continuous (quantitative)variables:

One variable denoted x ,is regarded as the predictor, explanatory, or independent variable.

The other variable denoted, y is regarded as the response, outcome, or dependent variable.

Here,the feature like the day of the week, holiday etc are the independent variables while the price of the flight is a dependent variable.

2.Random Forest Regression

The random forest model is a type of additive model that makes predictions by combining decisions from a sequence of base models. More formally we can write this class of models as:

$$G(x) = f_0(x) + f_1(x) + f_2(x) + \dots$$

where the final model G is the sum of simple base models f_i . Here, each base classifier is a simple decision tree. This broad technique of using models to obtain better predictive performance is called model ensembling. In random forests, all the base models are constructed independently using a different subsample of the data.

The model requires various parameters to be entered. The parameters and their chosen values as follows:

1. n_estimators: integer, optional (default =10)

This tells the number of trees in the forest. We choose 100 as value for this parameter. We conducted various test and concluded this value to be most optimal.

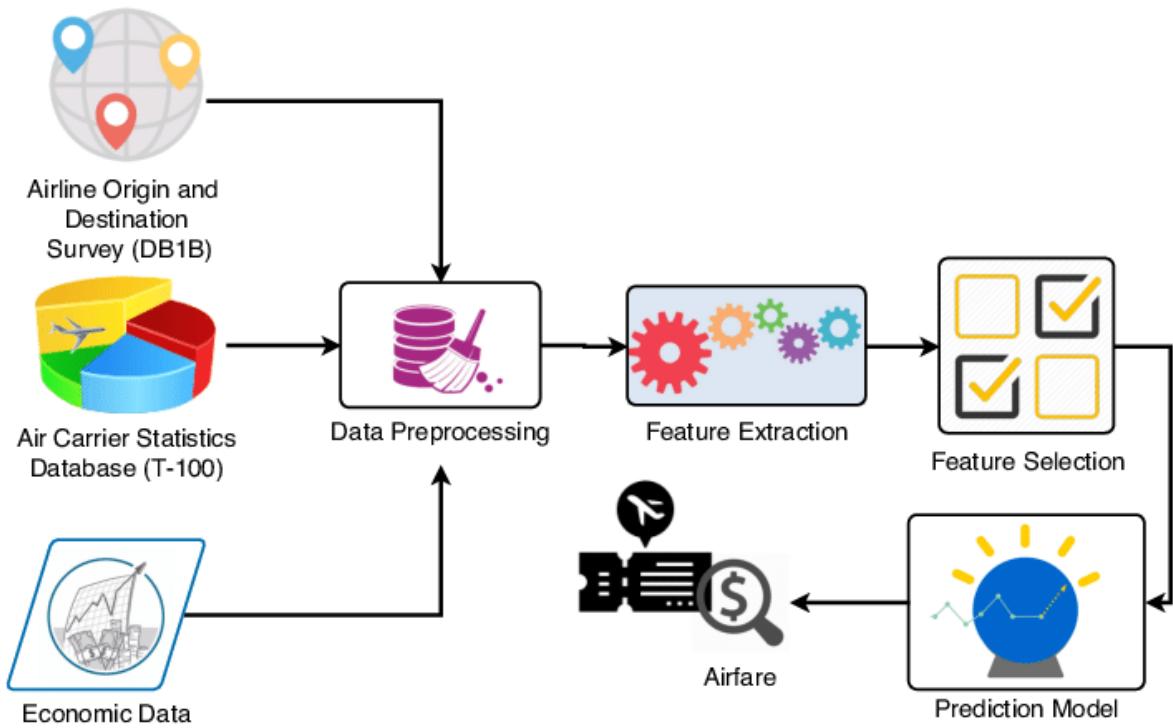
2. Max_depth: integer or none, optional(default equal to none)

The maximum depth of the tree if null then nodes are expected until all leaves are pure or until all leaves contain less than min_sample_split samples.

Since the number of features is less than 10 we decided to keep the maximum depth of the trees numerically low. After analysing the rest results with various values for Max_depth ,we concluded three as the most optimal Max_depth.

3. random_state: int, RandomState instance or None, optional (default=None)

If int, random_state is the seed used by the random number generator ;if RandomState instance, random_state is the random number generator. If None, the random number generator is the RandomState instance used by np.random.



4 RESULT AND DISCUSSION

The main goal is to predict the fares of the flights based on different factors available in the provided dataset which has been achieved using this model .The score achieved is Random Forest Regressor score = 0.887399115599125

The proposed system is an online system which helps the user to go through the rates quoted by the travel agency. The System offers a number of clear benefits. They reduce document distribution costs, eliminate complex calculations, and save customers' time.

The model ensures following features:

- Ensure data accuracy
- Better service
- Minimum time needed for the various processing
- Any person across the world, having internet can access this service
- Help the customers prepare and organize its holidays and trips more efficiently.

5.Conclusion and Future Scope

5.1 Conclusion

Our model **Random Forest Regressor** performed well on the prediction with an **accuracy of 95% on the train data and about 80% on the test data**

The project entitled “Flight Prize Prediction” developed using Machine Learning and python is an attempt to computerize the different operations in travel agencies. The project is very flexible and secure; admin’s can incorporate new features and manage the modules of the system ass per requirements. It can be a web based software and then it can be accessed from anywhere with internet. Tourism is currently recognized as a global industry which is growing at a high rate..

5.2 Future Scope

- Any tourist agency can make use of it for saving customer costs or consulting them about the same.
- Tourism group can use it for managing their travel costs
- This application can easily be implemented under various situations.
- We can add new features as an when required
- Reusability of the application is also possible.

References:-

<https://www.geeksforgeeks.org/machine-learning/>

<https://www.geeksforgeeks.org/introduction-machine-learning/>

<https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/>

<https://www.altexsoft.com/blog/flight-price-predictor/>

<https://www.analyticsvidhya.com/blog/2021/06/flight-price-prediction-a-regression-analysis-using-lazy-prediction/>

<https://www.kaggle.com/vinayshaw/airfare-price-prediction/data>