FLIGHT_PRICE_PREDICTION

NAME OF THE PROJECT

Submitted by:

LODEM RAKESH

YOUR NAME

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

Google,Wikipedia and some educational websites

INTRODUCTION

- Business Problem Framing

  Describe the business problem and how this problem can be related to the real world.

  It is the real world problem that predicts the price of different airlines.

- Conceptual Background of the Domain Problem

  Describe the domain related concepts that you think will be useful for better understanding of the project.

  Some understanding about the planes and their operation

- Review of Literature

  This is a comprehensive summary of the research done on the topic. The review should enumerate, describe, summarize, evaluate and clarify the research done.

  Google

- Motivation for the Problem Undertaken

  Describe your objective behind to make this project, this domain and what is the motivation behind.
  To deeply understand the project and its domain related problem

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

- Data Sources and their formats

  What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

  Webscrapping the different websites related to the flights opertion

- Data Preprocessing Done

  What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

  Checking the null values

  Dropping the null values

  Imputing the null values

  Handling the categorical columns

- Data Inputs- Logic- Output Relationships

  Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

  They are non linear in nature

- State the set of assumptions (if any) related to the problem under consideration

  Here, you can describe any presumptions taken by you.

  No assumptions taken

- Hardware and Software Requirements and Tools Used

  Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

  Numpy package,pandas,seaborn and matplotlib..etc

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Describe the approaches you followed, both statistical and analytical, for solving of this problem.

  Null values identification

  Standard scaler

Mutual info classifier

Randomizedsearchcv

- Testing of Identified Approaches (Algorithms)

  Listing down all the algorithms used for the training and testing.

  RandomForestClassifier

  LogisticRegression

  decisiontreeRegressor

  gradientBoostingRegressor

  ada boost classifier

- Run and Evaluate selected models

  Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

- Key Metrics for success in solving problem under consideration

  What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

  Mean absolute error
  Mean squared error
  R2score

- Visualizations

  Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

Distplot

Boxplot

Pairplot

- Interpretation of the Results

  Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

  There seems to be a realation between no of stops and price

CONCLUSION

- Key Findings and Conclusions of the Study

  Describe the key findings, inferences, observations from the whole problem.

  There seems to be a realation between no of stops and price

- Learning Outcomes of the Study in respect of Data Science

  List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

  RandomForestRegressor works well with best r2score

- Limitations of this work and Scope for Future Work

There is a lot scope in this domain

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.