https://lh4.googleusercontent.com/nbLiGCv2aaBycn0plTAHIlWExAELP58huYs-r-tn1BM-FZUQPbR0_cr5ekE90n5scNUECHFa2BloPpFkFClIDcM4Lzyy_QIk_4azoEAg3OOUuwTbyLWUEDhA8CewkpMN-cwlMq6vVVmUXW-uafewS6E

HIGH LEVEL DOCUMENT

**Package Travel Purchase PREDICTION**

September 2, 2022

***Written by othmane zoubairi***

CONTENTS

**DOCUMENT VERSION CONTROL**

3 **ABSTRACT** 4 **1) INTRODUCTION** 5 **1.1 WHY THIS HIGH-LEVEL DESIGN DOCUMENT?** 5 **1.2 SCOPE** 5 **1.3 DEFINITIONS** 5 **2) GENERAL DESCRIPTION** 6 **2.1 PRODUCT PERSPECTIVE** 6 **2.2 PROBLEM STATEMENT** 6 **2.3 PROPOSED SOLUTION** 6 **2.4 TECHNICAL REQUIREMENTS** 6 **2.5 DATA REQUIREMENTS** 7 **2.6 TOOLS AND TECHNOLOGIES USED** 8 **2.7 CONSTRAINTS** 8 **3)** 9 **3.1 PROCESS FLOW** 9 **3.2 DEPLOYMENT PROCESS** 9 **3.3 EVENT LOG** 10 **3.4 PERFORMANCE** 10

**3.5 REUSABILITY** 10 **3.7 RESOURCE UTILIZATION**

11 **3.9 DEPLOYMENT**

11 **3.1O USER INTERFACE** 12

**4 Conclusion**

2 | P a g e

**DOCUMENT VERSION CONTROL**

|  |  |  |  |
| --- | --- | --- | --- |
| Date  Issued | Version | Description | Author |
| 13/07/2022 | 1.0.1 | Initial HLD – V1.0.1 | Othmane  zoubairi |
| 10/08/2022 | 1.0.2 | Updated HLD  – V1.0.2 | Othmane  zoubairi |
| 02/09/2022 | 1.0.3 | Updated HLD  – V1.0.3 | Othmane  zoubairi |
|  |  |  |  |

3 | P a g e

ABSTRACT

The Travel company wants to enable and establish a viable business model to expand the customer base. One of the ways to expand the customer base is to introduce a new offering of packages.

With the help of Data Science and Machine learning  technology, I developed an application, which allows a Travel Company to determine the probability Of Package Purchase in just a  few seconds.

4 | P a g e

  1) Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-level document (HLD) is to describe  the design of the project in detail which can be used as a  reference manual.

The HLD will:

• Present all the design aspects and define them in detail. • Describe the user interface being implemented. • Describe the software interfaces.

• Describe the performance requirements.

• Include design features and the architecture of the project.

1.2 Scope

The HLD document presents the entire structure of the project  in parts, such as the data ingestion, data pre-processing,  solution development, and the deployment part along with their  respective architectures. This uses non-technical to mild  technical terms which should be understandable to the

administrators of the system.

1.3 Definitions

|  |  |
| --- | --- |
| Term | Description |
| EDA | Exploratory Data Analysis |
| IDE | Integrated Development Environment |

5 | P a g e

2) General Description

2.1 Product Perspective

The Travel Package Purchase Probability predictor is a machine  learning based on the use of GridSearchCV Bagging Classifier model which will help us  to predict the probability Of Package Purchase based on the attributes of  the customer.

2.2 Problem Statement

The Travel company wants to enable and establish a viable business model to expand the customer base. One of the ways to expand the customer base is to introduce a new offering of packages. Base on Data Company we need to analyize

The Data of Clients to find who is able to sell new Travel Package.

2.3 Proposed Solution

The solution proposed here is a web application, which takes  the details of the customer and those details will be taken by a  machine learning model in the backend, which will then predict the probability of Package Purchase and display it on the front-end page of  the user.

2.4 Technical Requirements

I used python version 3.7 with some important libraries to develop a machine learning model, which accurately predicts  the probability of Travel Package Purchase.

Then, the model is used as a back-end software for a front-end  web application which can be used by the users.

6 | P a g e 

 2.5 Data Requirements

For training and testing the model, I used the public set  available in Ineuron intership, “

URL – <https://raw.githubusercontent.com/manotti22/Travel-Package-Prediction-Projekt/main/TourismData.csv>

Dataset Attributes:

Customer details:[¶](https://www.kaggle.com/code/dscodingp19/travel-package-purchase-prediction" \l "Customer-details:" \t "_self)

1. CustomerID: Unique customer ID
2. Age: Age of customer
3. TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
4. CityTier: City tier depends on the development of a city, population, facilities, and living standards.
5. Occupation: Occupation of customer
6. Gender: Gender of customer
7. NumberOfPersonVisiting: Total number of persons planning to take the trip with the customer
8. PreferredPropertyStar: Preferred hotel property rating by customer
9. MaritalStatus: Marital status of customer
10. NumberOfTrips: Average number of trips in a year by customer
11. Passport: The customer has a passport or not (0: No, 1: Yes)
12. OwnCar: Whether the customers own a car or not (0: No, 1: Yes)
13. NumberOfChildrenVisiting: Total number of children with age less than 5 planning to take the trip with the customer
14. Designation: Designation of the customer in the current organization
15. MonthlyIncome: Gross monthly income of the customer

Label Column ;

1. ProdTaken: Whether the customer has purchased a package or not (0: No, 1: Yes)

2.6 Tools And Technologies Used



• Jupyter notebook is used for EDA and experimentation  with various ML algorithms with the help of pandas,  numpy, matplotlib, seaborn, sklearn   libraries.

• Jupyter was also used for the development and  deployment of the solution with logging. Used python  version 3.7 and libraries include logging, pandas, numpy,  scikit learn, flask, and HTML

* Github is used as a version control system. ‘
* Using Git to have acces to github to make uptade evry time you needed
* Deployed on the web using Gunicorn and Heroku.

2.7 Constraints

The Concrete Compressive Strength Prediction system must be  user-friendly, errors free and users should not be required to  know about any of the workings.

8 | P a g e

 3 Event Log

In this project, I used the “logging” library in both the  development and deployment stages, which keeps logging the  events at every step into the “ Travel\_logs” files. One of the advantages  of event logging is, it makes debugging much easier, we can  directly go to that specific line of code, which has errors.

3.1 Performance

The ML-based Probability Of Purchase Predictor application is  used for predicting the Probability Of Package Purchase based on various  attributes of the customer. So, it should be as accurate as  possible, so that it will not mislead the Travel Company.

Model retraining is very important to keep it relevant in order  to keep the model dynamic to changing times and customer behaviour.

3.2 Reusability

The code written and the components used have the ability to  be reused without any problem.

3.3 Application Compatibility

The different components or modules of this project use python version 3.7 as their interface between them. Each component has its own task to perform, and it is the job of the python version to ensure proper transfer  of the information.

3.4 Resource utilization

In this project, any task may use all the processing power  available in the system, until it is accomplished.

3.5 Deployment

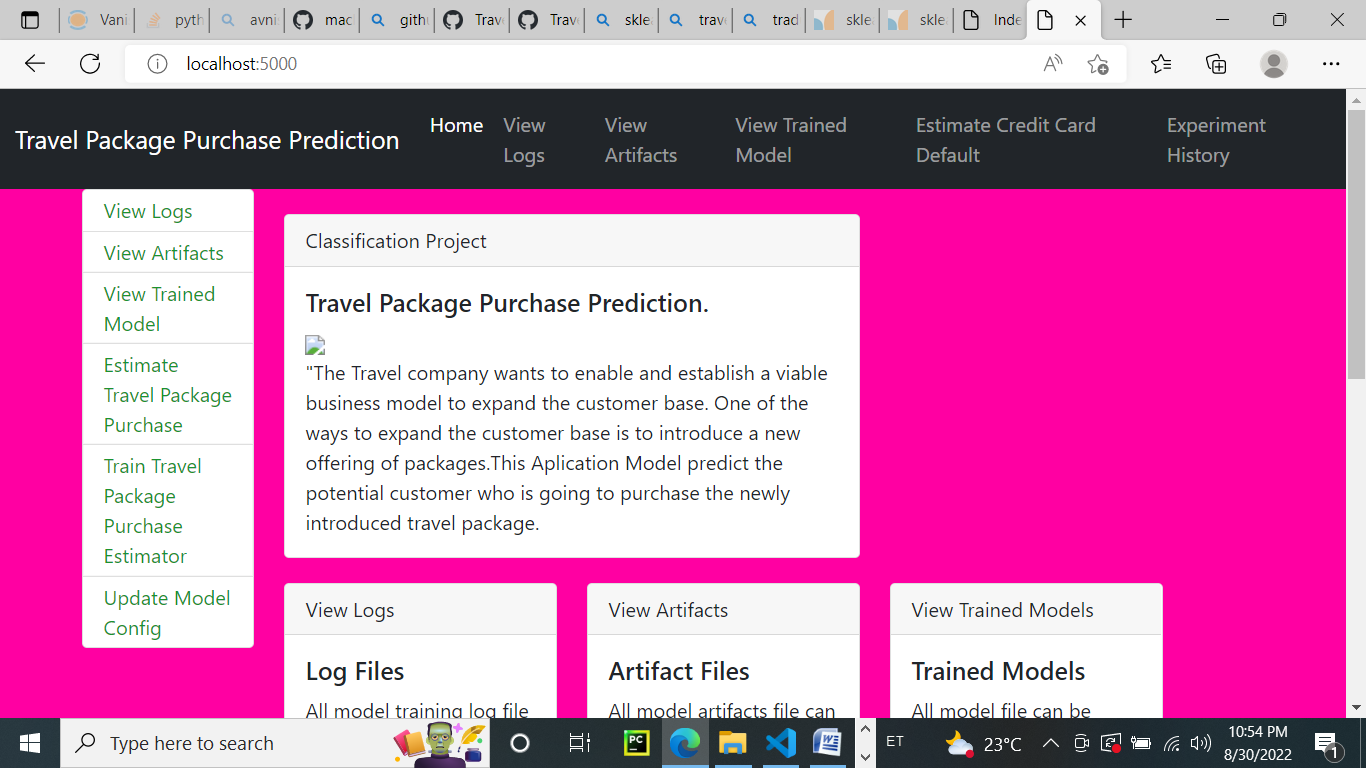
I deployed the application on the web using Heroku

\*URL - https://- travelpackage-predherokuapp.com/

\*Deployemnt dosen work in Heroku Free Version if the size data more than 500 MB in free version

from heroku

3.6 User Interface



4 Conclusion

The company can run the model to achieve desired performance levels for new data, also to offer better packages to customers.

Young and single people are more likely to buy the offered packages.

Age and Income have a correlation and we see that higher age groups and higher Monthly Income groups lean towards the expensive packages.

**Recommendations**

The marketing team can curate the individual packages to the specific business designation

The marketing team can create product and customer segment specific sale pitch to reduce the DurationOfPitch.

The WELLNESS TOURISM PACKAGE should be curated considering the features of existing packages that customers have purchased.

The company can run various campaigns and offers for customers with family to increase sales.

The data shows customers with passport has higher buying ratio and business can curate international packages for such customers.

Specific packages can be created for different income groups.