Hight Level Design (HLD)

### Travel Package Purchase Prediction

# Revision Number: 2.0

Last date of Revision: 26/03/2023

## Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 14/12/2022 | 1 | Initial HLD - V1.0 | Sonali D. |
| 26/03/2022 | 2 | Updated KPI - V1.1 | Sonali D. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

#### 

#### Content

Document version control………………………………………………………………………………………………….2

Abstract………………………………………………………………………………………………………………………………

1. Introduction…………………………………………………………………………………………………………………….5
   1. What is High-Level Design document………………………………………………………………….5
   2. Scope………………………………………………………………………………………………………………….5
2. General Description……………………………………………………………………………………………………….6

2.1 Product Perspective……………………………………………………………………………………………6

* 1. Problem Statement…………………………………………………………………………………………….
  2. Proposed Solution…………………………………………………………………………………………….
  3. Further Improvements………………………………………………………………………………………
  4. Technical Requirements……………………………………………………………………………………
  5. Data Requirements…………………………………………………………………………………………..
  6. Tools Used……………………………………………………………………………………………………….
  7. Constraints………………………………………………………………………………………………………

1. Design Detail……………………………………………………………………………………………………………….

3.1 Process Flow……………………………………………………………………………………………………

* + 1. Model Training and Evaluation………………………………………………………….
  1. Performance……………………………………………………………………………………………………
  2. Reusability……………………………………………………………………………………………………….
  3. Application Compatibility……………………………………………………………………………….
  4. Resource Utilization………………………………………………………………………………………
  5. Deployment………………………………………………………………………………………………….

1. Dashboards…………………………………………………………………………………………………………………
   1. KPI (Key Performance Indicators) ………………………………………………………………………
2. Conclusion…………………………………………………………………………………………………………………

# Abstract

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. Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking at the data of the last year, we observed that 18.9% of the customers purchased the packages. However, the marketing cost was quite high because customers were contacted at random without looking at the available information. The company wants to know that the customer will buy the package or not. However, this time company wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

We need to analyse the customers' data and information to provide recommendations to the Marketing Team and also build a model to predict the potential customer who is going to purchase the package. Tourism is one of the most rapidly growing global industries and tourism forecasting is becoming an increasingly important activity in planning and managing the industry. Because of high fluctuations of tourism demand, accurate predictions of purchase of travel packages are of high importance for tourism organizations. Here I have analysed the trend using the data.

I have predicted whether the customer will purchase the travel or not.

# **Introduction**

## 1.1 Why this High-Level Design Document?

## The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

## The HLD will:

## Present all of the design aspects and define them in detail

## Describe the user interface being implemented

## Describe the hardware and software interfaces

## Describe the performance requirements

## Include design features and the architecture of the project

## List and describe the non-functional attributes like:

* Security
* Reliability
* Maintainability
* Portability
* Reusability
* Application compatibility
* Resource utilization
* Serviceability

# **Scope**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

## 2 General Description

## 2.1 Product Perspective

The Travel Package Purchase Prediction predict whether the customer will purchase the travel or not which will help us to detect the anomalies in the travel industry and take the necessary action.

## 2.2 Problem statement

1. To predict which customer is more likely to purchase the travel package
2. Which variables are most significant.
3. Which segment of customers should be targeted more.

## 2.3 Proposed Solution

The data that is used has gone through multiple stages like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing.

Different machine learning algorithms that’s best fit for the above case.

## 2.4 Further Improvements

The Travel Package Purchase Prediction Data model has gone through various regression techniques to improve the accuracy but other regressions or techniques can be applied to check further.

## 2.5 Technical Requirements

We have created our Data model using almost 5000 rows of data, so to improve the model more data is needed. The more the data the accurate the model.

## Data Requirements

Data requirement completely depend on our problem statement.

Customer details:

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

Customer interaction data:

* PitchSatisfactionScore: Sales pitch satisfaction score
* ProductPitched: Product pitched by the salesperson
* NumberOfFollowups: Total number of follow-ups has been done by the salesperson after the sales pitch
* DurationOfPitch: Duration of the pitch by a salesperson to the customer

## Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn are used to build the whole model.



* PyCharm is used as IDE.
* For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
* Tableau/Power Bl is used for dashboard creation.
* data analysis and associated manipulation of tabular data in data frames.
* GitHub is used for storing.

## Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

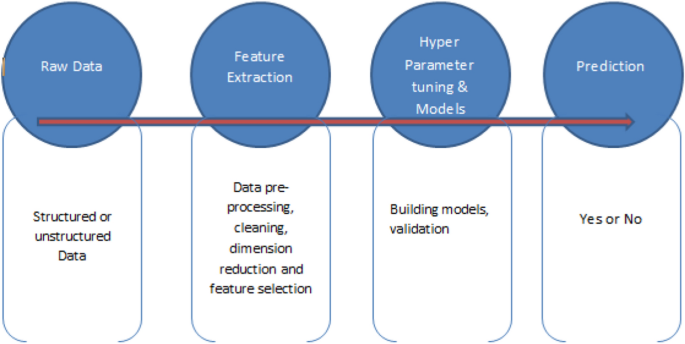
## Constraints

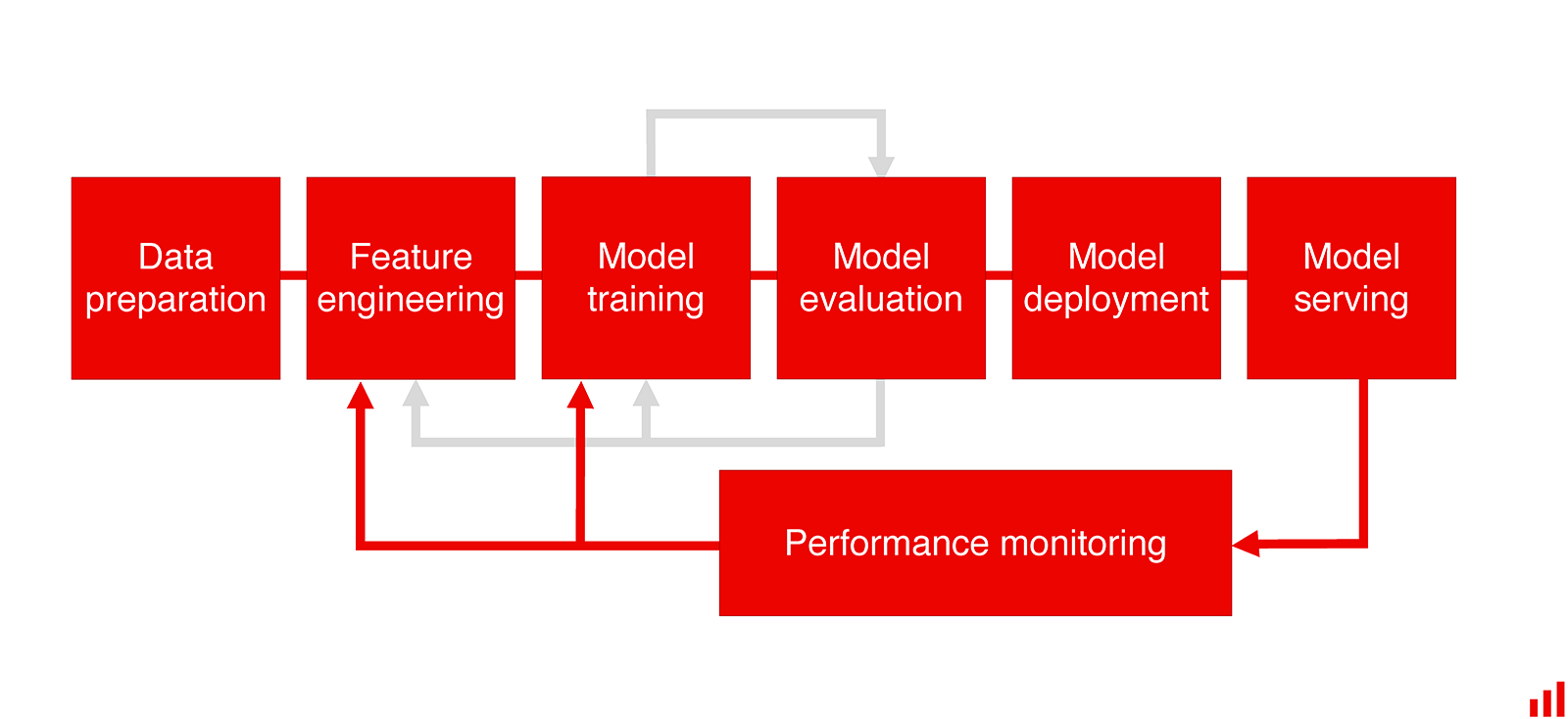
The challenge here would be how to increase the success ratio for customers taking the packages. There is a larger portion of people who do not have passport. Finding which variables would impact the success ratio is crucial. There was a lot of data available but needed to figure out how to make sense of it and identify actionable insights.

## Design Details

## Process Flow

Below is the process flow diagram is as shown below.

Proposed methodology

3.1.1 Model Training and Evaluation

## Performance

* + - Tuned Decision Tree gives a more generalized model.
    - XGboost seems to overfit. To get more generalized model, we can be look into tuning XGboost with different parameters.
    - we can also tune stacking with different weak learners which can help improve the performance and get a generalized model.

## Reusability

The code written and the components used should have the ability to be reused with no problems.

## Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

## Resource Utilization

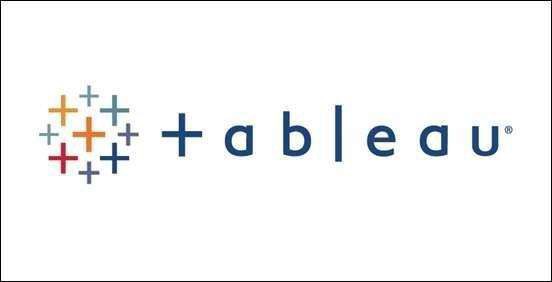
When any task is performed, it will likely use all the processing power available until that function is finished.

## Deployment



## Dashboards

Dashboards will be implemented to display and indicate certain KPls and relevant indicators for the unveiled problems that if not addressed in time could cause catastrophes of unimaginable impact.



As and when, the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors.

## KPls (Key Performance Indicators)

* Model is showing the most important feature from the dataset that can be considered to make business driven decisions.
* Helping to make accurate decisions and business plans
* Model is helping to detect the customers that needs to be more focused on.
* Model is helping to understand who are most likely to purchase the package.
* The model implemented will help to understand specially on which customers to be focused more and where we need better ideas to be implemented instead of applying the same methodology to all customers. This will increase revenue and reduce marketing costs.

## Conclusion

* + - Looking at feature importance Designation, Passport, Tier city, marital status, occupation are most important features. Income can also be looked into, few of other models have given higher importance to Income as well.
    - Gender, number of children visiting, having a car seemed to be insignificant.

## References

1. <https://www.kaggle.com/code/yogidsba/travelpackageprediction-ensemble-techniques>
2. [scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation](https://scikit-learn.org/stable/index.html)