# Architecture Design

# Travel Package Purchase Prediction

**Revision No: 1.0**
**Last Date of Revision: 22/07/23**

**Manasi Khillare**
**Naresh K**

# Document Version Control

| Date Issued | Version | Description | Author |
|---|---|---|---|
| 22/07/23 | 1.0 | First Version of Complete Architecture Design | Manasi Khillare<br>Naresh K |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Contents

# Abstract

Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build models and employ algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. These models can be applied in different areas and trained to match the expectations of management so that accurate steps can be taken to achieve the organization's target. In this project, we will estimate the Travel Package Purchase on the basis of personal health information. Taking various aspects of a dataset collected from people, and the methodology followed for building a predictive model.

# 1.Introduction

### 1.1 What is Architecture design document?

Any software needs the architectural design to represents the design of software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures. Each style will describe a system category that consists of:
A set of components (e.g.: a database, computational modules) that will perform a function required by the system. The set of connectors will help in coordination, communication, and cooperation between the components. Conditions that how components can be integrated to form the system. Semantic models that help the designer to understand the overall properties of the system.

### 1.2 Scope

Architecture Design Document (ADD) is an architecture design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

### 1.3 Constraints

We only predict the whether the customer purchases the package or not

# 2.Technical Specification

## 2.1 Data dictionary

- Customer ID - Unique customer ID Prod Taken: Whether the customer has purchased a package or not (0: No, 1: Yes)
- Age - Age of customer
- Type of Contact - How customer was contacted (Company Invited or Self Inquiry)
- City Tier - 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
- Number of Persons Visiting - Total number of persons planning to take the trip with the customer
- Preferred Property Star - Preferred hotel property rating by customer
- Marital Status - Marital status of customer
- Number of Trips - Average number of trips in a year by customer
- Passport - The customer has a passport or not (0: No, 1: Yes)
- Own Car - Whether the customers own a car or not (0: No, 1: Yes)
- Number of Children Visiting - 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
- Monthly Income - Gross monthly income of the customer
- Pitch Satisfaction Score - Sales pitch satisfaction score
- Product Pitched - Product pitched by the salesperson(type of package chosen)
- Number of Follow ups - Total number of follow-ups has been done by the salesperson after the sales pitch
- Duration of Pitch - Duration of the pitch by a salesperson to the customer(call duration)

## Data Pre-Processing

Data preprocessing is a crucial step in any machine learning project. It involves transforming, cleaning, and preparing the raw data for further processing and modeling. Data preprocessing aims to improve the quality and usability of the data by removing noise, outliers, missing values, inconsistencies and irrelevant features. It also aims to enhance the data by normalizing, scaling, and encoding.

It can have a significant impact on the performance and accuracy of the subsequent machine learning models. Therefore, it is important to apply appropriate and effective data preprocessing techniques based on the characteristics and requirements of the data and the problem.

## 2.2 Logging

- We should be able to log every activity done by the user.
- The system identifies at which step logging require.
- The system should be able to log each and every system flow.
- The system should not be hung even after using so much logging.
- Logging is just because we can easily debug issuing so logging is mandatory to do.

## 2.3 Deployment

For the hosting of the project, we use ____________

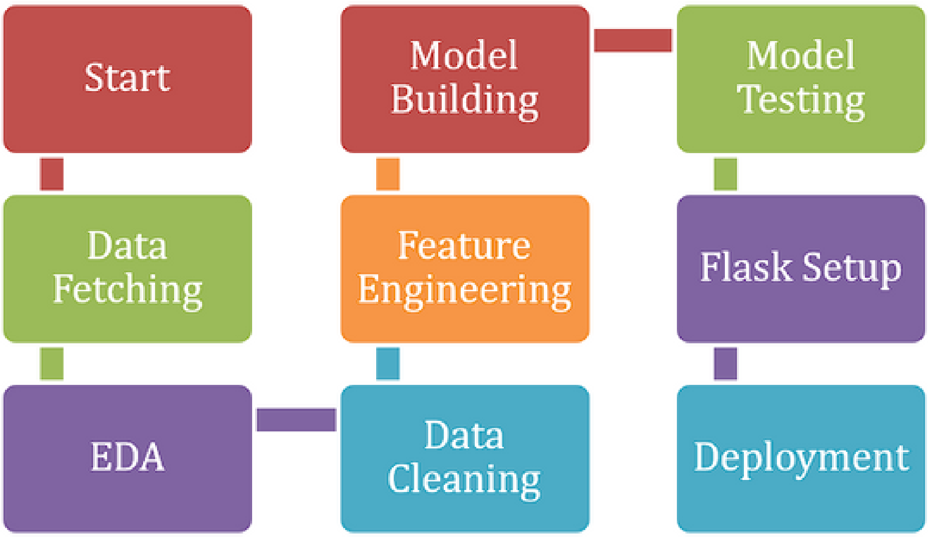
# 3. Technology Stack

| | |
|---|---|
| Front End | HTML/CSS |
| Backend | Python/ Flask |
| Deployment | Heroku |

## 3.1 Proposed Solution

We will use performed EDA to find the important relation between different attributes and will use a machine-learning algorithm to estimate the 'Package Purchase Prediction'. The client will fill in the required feature as input and get results through the web application. The system will get features and it will be passed into the backend where the features will be validated and pre-processed and then it will be passed to a best performing machine learning model to predict the final outcome.

# 5.Architecture



## 5.1 Data Gathering

Data source:  iNeuron dataset
Train and Test data are stored in xlsx format.

## 5.2 Raw Data Validation

Raw data validation is the process of ensuring the accuracy and quality of raw data, which is data that has been collected from one or multiple sources but is still in its initial, unaltered state. It involves checking the datatype, whether the data follows a certain format, logically consistent with other data, uniqueness check. Raw data validation can improve the reliability and usability of the data for further analysis or processing.

## 5.3 Exploratory Data Analysis

Univariate analysis using descriptive statistics and Bivariate analysis to visualize relationship between dependent and independent features and also relationship between independent features has been performed to get insights about the data.

## 5.4 Feature Engineering

Outliers have been treated to ensure the data is normally distributed. One hot encoding has been performed to convert the categorical features into numerical features. Best features were selected (using Chi-Square test).

## 5.5 Model Building

The model is saved using the pickle library in ‘pickle’ format.
Model building phase is the process of creating and training a machine learning model. Various models like Logistic Regression, Decision Tress, XG Boost were tested and cross validation has been performed to ensure that the model generalizes for the unseen data. Hyperparameter tuning has also been performed to get the best accuracy.

## 5.6 Model Saving

The model is saved using the pickle library in ‘pickle’ format.

## 5.7 Flask Setup for Web Application

After saving the model, the API building process started using Flask. Web application creation was created in Flask for testing purposes. Whatever user will enter the data and then that data will be extracted by the model to estimate the premium of insurance, this is performed in this stage

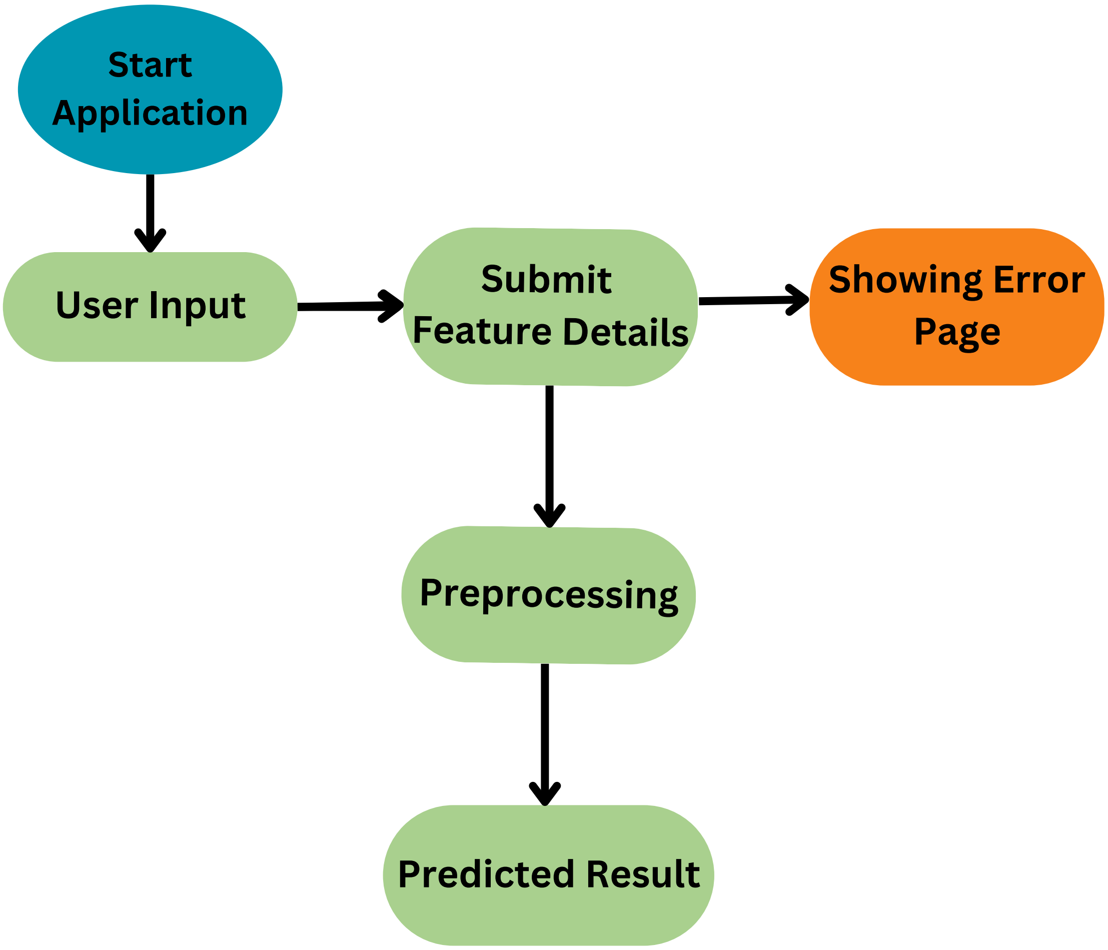
## 5.8 GitHub

The whole project directory will be pushed into the GitHub repository.

## 5.9 Deployment

The project was deployed from GitHub into the Heroku platform

## 6.User Input / Output Workflow:


Start Application
User Input
Submit Feature Details
Showing Error Page
Preprocessing
Predicted Result