Insurance Premium Prediction

High Level Design

Domain: Machine Learning

Author: Alok Kumar

Date: 16.04.2024

Document Version Control

Date issued	Version	Description	Author
Nov 25th 2023	0.0.1	First Draft	Alok Kumar

Delhi, India

+91 6205197087

kalok0575@gmail.com

CONTENTS

ABSTRACT

Health Insurance is one of the first and foremost aspects of financial and mental freedom. Health is a family's biggest asset and to keep it protected is one's foremost duty. In our daily lives, we go through various aspects of struggle; however, health is one factor which should never be ignored. Hence, we should be proactive in protecting and safeguarding our and the loved ones' future.

INTRODUCTION

WHAT IS HIGH-LEVEL DESIGN DOCUMENT?

The goal of this HLD or a high-level design 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 design aspects and define them in detail.
- Describe all user interfaces being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and architecture of the project.
- List and describe the non-functional attributes such as security, reliability, maintainability, portability, reusability, application compatibility, resource utilization, serviceability.

SCOPE

The HLD documentation presents the structure of the system, such as 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.

GENERAL DESCRIPTION

DEFINITIONS

Term	Description
IPP	Insurance Premium Predictor.
Database	Collection of the Information.
Cloud	A data center full of services connected to the internet performing service.
IDE	Integrated Development Environment.

PRODUCT DESCRIPTION

IPP is a Machine Learning based regression model which helps us to predict premium expenses of a person.

PROBLEM STATEMENT

The goal of this project is to give people an estimate of how much they need based on their individual health situation. After that, customers can work with any health insurance carrier and its plans and perks while keeping the projected cost from our study in mind. This can assist a person in concentrating on the health side of an insurance policy rather than the ineffective part.

PROPOSED SOLUTION

Using all the standard techniques used in the life cycle of a Data Science project starting from Data Exploration, Data Cleaning, Feature Engineering, Model Selection, Model Building and Model Testing and also building a frontend where a user can fill their information in the form input and get the output instantly.

DATA REQUIREMENTS

Data requirement completely depend on our problem statement. We need the dataset from Insurance Company. Required dataset should contain the following features:

- Age of the client.
- Whether the client is smoker or no.
- The region of the client.

3

- Number of children the client have.
- The body mass index of the client.
- Gender of the client.

These are the required parameters to feed into model.

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Flask and a few other libraries were used to build the whole model.

- For visualization tasks, Matplotlib and Seaborn were used.
- Flask is used for building the web application and server to run the code.
- GitHub is used as version control system.
- NumPy and Pandas were used to clean and interpret data.
- Sckikit-learn was used to cross validate and compare different models.

HARDWARE REQUIREMENTS

- Windows Server, Linux, or any operating system that can run as a webserver, capable of delivering HTML content.
- Minimum 1.10 GHz processor or equivalent.
- Between 1-2 GB of free storage.
- Minimum 512 MB of RAM.
- 3 GB of hard-disk space.

The main objective of this project is to implement the use case as previously mentioned (2.3 problem statement) for new dataset that comes through the form. It is assumed that all aspects of this project have the ability to work together as the designer is expecting and also the data on which our model is trained is as correct as possible.
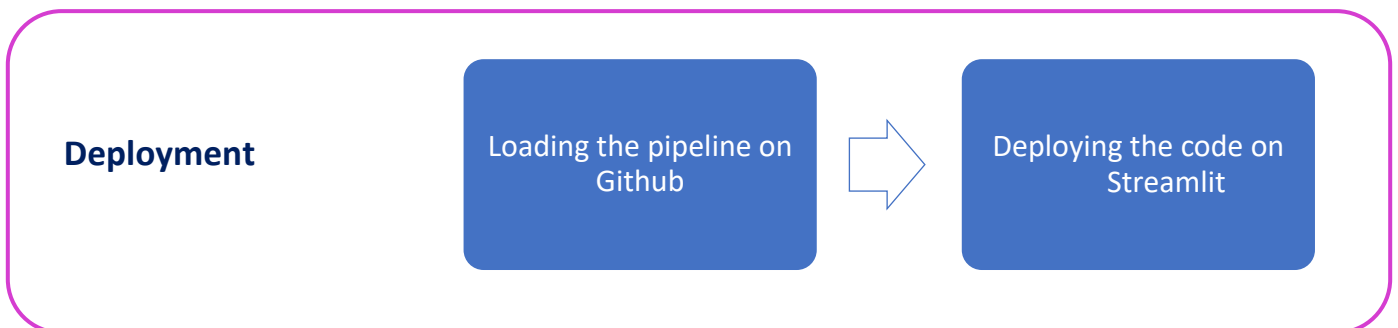
DESIGN DETAILS

PROCESS FLOW

For accomplishment of the task, we will use a trained Machine Learning model. The process flow diagram is shown below:
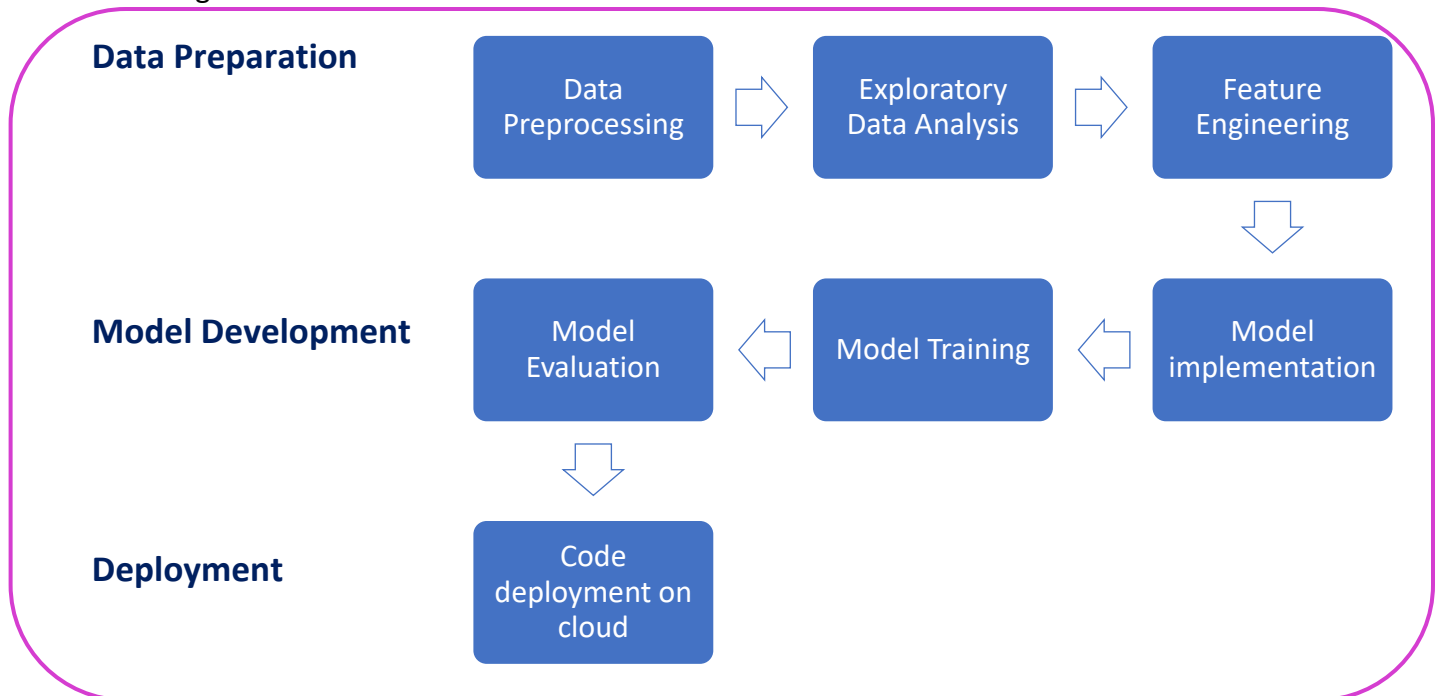
DEPLOYMENT DIAGRAM

EVENT LOG

The system should log every event so that the user will know what process is running internally. Initial step-by-step description:

1. The system identifies at what level logging is required.

2. The system should be able to log each and every system flow.

3. You can choose logging method. You can choose database logging/ File logging as well.

4. System should not hang even after so many loggings. Logging just because we can easily debug issues, so logging is mandatory to do.

ERROR HANDLING

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

PERFORMANCE

The IPP tool is used to predict the expenses of the person on the insurance premium based on the health condition and other parameters. Also, model retraining is very important to further enhance its performance.

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 Python to ensure proper transfer of information.

RESOURCE UTILIZATION

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

KPIS (KEY PERFORMANCE INDICATORS)

- Key Performance Indicators of IPP
- Latency or the amount of time the application takes to display results for some specific input.
- The processing power our application takes to run
- The memory and RAM our application takes to run on a web server.

CONCLUSION

All in all, overall project architecture, design details, used technologies and performance were explained in detail. The IPP will give the people the idea of the expenses they have to make on their health insurance premium.