HIGH LEVEL DESIGN

INSURANCE PREMIUM PREDICTION

Document Version Control

Date Issued	Version	Description	Author
5/9/24	V1.0	HLD- V1.0	Madhuri Chandanbatwe

Abstract

In this project,we analyze the personal health data to predict insurance premium of individuals. Predictive insurance is an advanced type of analysis that allows insurance companies to make forecasts using their historical data, combining statistical models, data mining techniques, and machine learning.

Machine learning can significantly enhance the prediction of insurance premiums through various techniques and methodologies through data analysis, selecting the most relevant features that contribute to premium calculations and predictive modelling.In this project we mostly focuss on building a premium prediction application that would help the insurance companies predict premiums for their customers.

1.0 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level document is to add necessary details to current project description to represent a suitable model for coding. This document is used as a reference manual for how the model interact at a high-level.

The HLD will

- Presents all 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 feature and the architecture of the project.

1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture, and technology architecture. The HLD uses non-technical to middle-technical terms which should be understandable to the administrators of the system.

2.1 Product Perspective

The Insurance premium estimation is a machine learning based predictive model which will help us to predict the premium of the person for health insurance.

2.2 Problem Statement

To predict health insurance premium of the person with the help of his historic data.

2.3 Proposed Solution

Machine learning models are trained with the given data and then the model gives predictions on unknown data.The problem comes under regression problem since the data is labelled.Effects of various features on the target label is observed and appropriate measures are proposed for the solution. By leveraging these capabilities, insurance companies can enhance their pricing strategies, improve customer satisfaction, and maintain competitiveness in the market.

2.4 Technical Requirements

Good internet connection

Cloud storage to store data like mongodb or AWS S3 bucket if data is large or if unstructured.

Local machine where the project was completed was 32 GB RAM on a Windows system on Visual Studio and Jupyter notebook.The application is compatible on Linux and MAC systems too.

2.5 Data Requirements

Data requirements completely depends on out problem statement.Data was available in .csv format.Data was downloaded from Kaggle data and was mostly of 14 GB.Data consisted of categorical as well as numerical columns.Each feature was treated according to the needs while data preparation tasks.Relevant transformations on the data was done to prepare the data for training.

2.6 Tools Used

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

2.6.1 Pandas is an open-source Python package that is widely used for data analysis and machine learning tasks.

2.6.2 NumPy is most commonly used package for scientific computing in Python.

2.6.3 Matplotlib,seaborn libraries are open-source data visualization library used to create interactive and quality charts/graphs.

2.6.4 Scikit-learn is used for a machine learning.

2.6.5 Flask is used to build API.

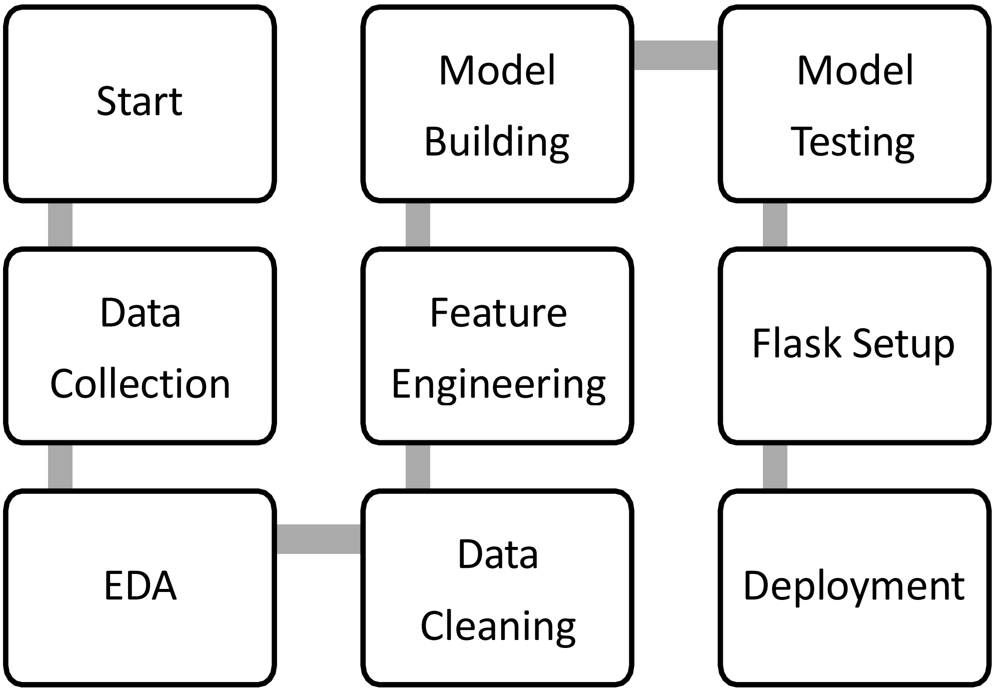
2.6.6 VS Code is used as IDE (Integrated Development Environment)

2.6.7 GitHub is used as version control system.

2.6.8 Front end development is done using HTML/CSS.

2.6.9 AWS is used for deployment of the model.

2.7 Process Flow



2.8 Event logging

The project logs every event successfully giving a clear idea about the errors and exceptions without disturbing the flow of code.

2.9 Reusability

The code is written in modular fashion enabling the reusability of code.The entire application is API oriented.

3.0 Deployment

Application can be deployed in any cloud platform like AWS,Heroku,Google Cloud or Azure whichever is suitable.Here we have used AWS for deployment purpose.Also the application is accessible through an API from any web browser.

3.1 Conclusion

The application was successfully able to predict the insurance premiums of the customers based on their health data.Further improvements in this application can be done by availability of more health related features of an indivisual that would help better predict the premiums more accurately.Here data being limited with few characteristic features available we were successfully able to create an application to predict the premiums of indivisuals.This would better help the insurance companies launch new health insurance schemes and offer indivisuals better health offers.