Architecture Design
Insurance Premium Prediction

Revision 1.0.1

Date: 07/01/2023

Description: Initial revision HLD

Document Version Control

- Change Information:

Date	Revision	Description	Author
08/01/2023	1.0.1	Initial revision (1.0.1)	Anand Chavan

- Reviewers:

Date	Revision	comments	Reviewer
08/01/2023	1.0.1	Initial revision (1.0.1)	Anand Chavan

- Approval status

Date	Revision	comments	Reviewer

Contents

1. Abstract

2. Introduction

2.1 Why architecture Design Document?

2.2 Scope

2.3 Definitions

3. Technical specification

3.1 Dataset overview

3.2 Predicting the Insurance premium

3.3 Logging

4. Technology Stack

5. Proposed Solution

6. Workflow 8

7. Key Performance indicators (KPI)

1. Abstract

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. I am considering variables as age, sex, BMI, number of children, smoking habits and living region to predict the premium. This can assist a person in concentrating on the health side of an insurance policy rather than the ineffective part.

2. Introduction

2.1 Why this architecture Design Document?

The goal of architecture Design Document is to give an internal logical design of the actual program code for the Insurance Premium Prediction System. architecture Design Document describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

2.2 Scope

Low-level design (LLD) is a component level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then defined during data design work.

2.3 Definitions

Term	Description
EDA	Exploratory Data Analysis
IDE	Integrated Development Environment
PaaS	Platform as a Service

3. Technical Specifications

3.1 Dataset Overview

For training and testing the model, I used the public data set available in Kaggle, “Insurance Premium Prediction” by nursnaaz

URL: https://www.kaggle.com/noordeen/insurance-premium prediction

Following is the data dictionary:

Name	Data Type	Description
Age	Integer	Input variable
Sex	String	Input variable
BMI	Decimal	Input variable
Children	Integer	Input variable
Smoker	String	Input variable
Region	String	Input variable
Expenses	Decimal	Output variable

3.2 Predicting the Insurance Premium

• The web application must be loaded properly for the users without any technical glitches like server timeouts.

• It must display the input fields and the “Predict” button to the users who accessed the application and allow the user to enter the values with respect to the personal information.

• The user gives the required information.

• Then the application should be able to predict the insurance premium based on the information given by the user.

3.3 Logging

We should be able to log every activity done by the user.
• The system should be able to log every step in the program flow.

• System should not be hung even after using so many loggings

• Logging makes debugging much easier, like we can directly go to that specific line of code, having bugs.

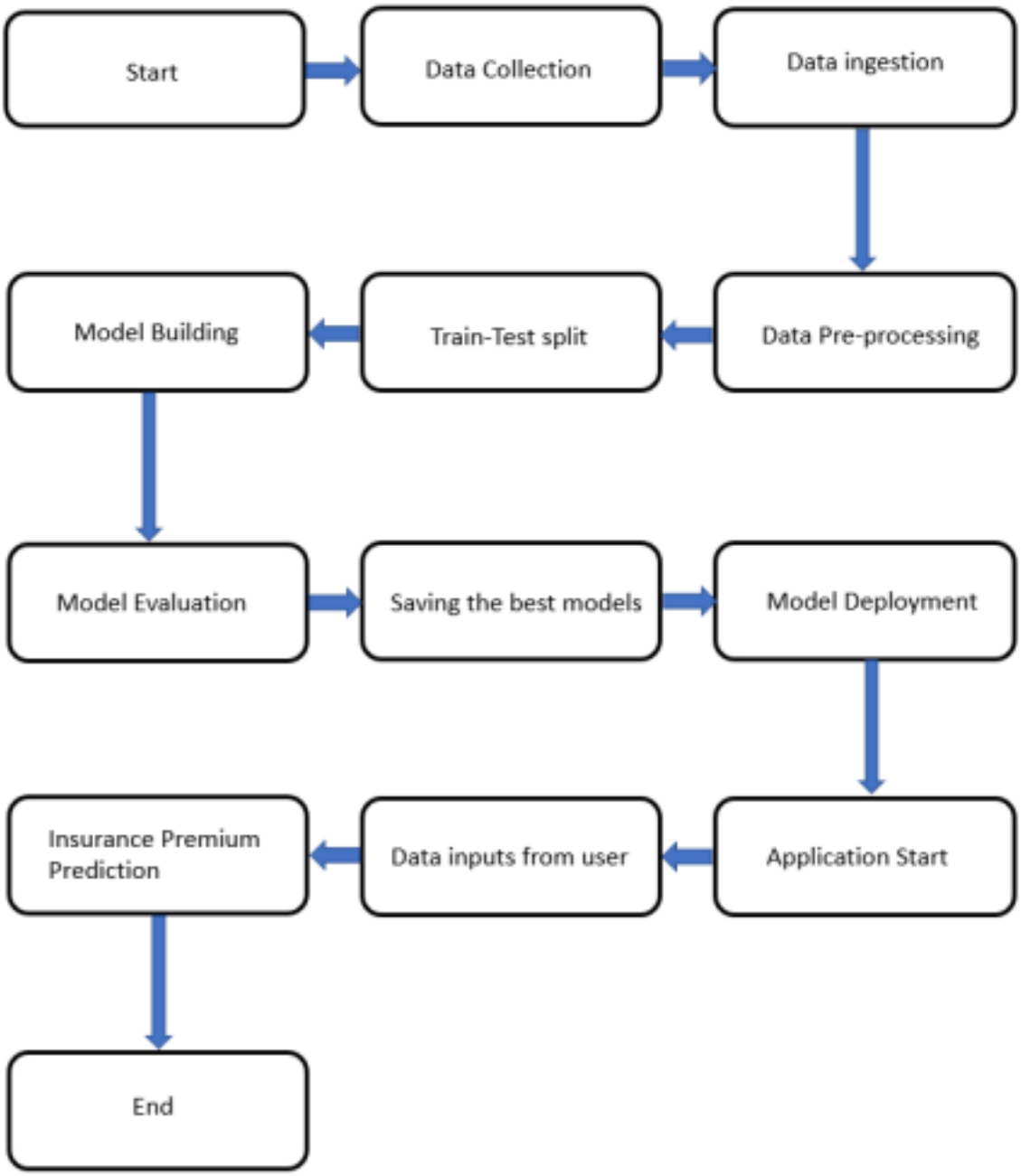
4. Technology Stack

Front-End	**Streamlit**
Back-End	**Python version 3.7**
Deployment	**Render**

5. Proposed solution

The solution proposed here is a web application, which takes the details of the personal information which contributes to insurance premium and those details will be taken by an Xgboost regressor model in the backend, which predicts the premium in dollars and displays in the front-end page to the user.

6. Workflow


Start
Data Collection
Data ingestion
Data Pre-processing
Train-Test split
Model Building
Model Evaluation
Saving the best models
Model Deployment
Application Start
Data inputs from user
Insurance Premium Prediction
End

7. Key Performance indicators (KPI)

• Time and workload reduction using the regression models. • Comparison of the R2 scores and the Adjusted R2 scores of the model on both the training and the testing data.

• Comparison of the RMSE scores of the model on both the training and the testing data.

• Feature importance using random forest regressor:

	Varname	Imp
4	yes	0.510938
0	age	0.411475
1	bmi	0.039646
2	children	0.032593
3	male	0.002795
6	region_southeast	0.000984
7	region_southwest	0.000957
5	region_northwest	0.000611