ARCHITECTURE

Insurance Premium Prediction

Last date of revision: 09/15/2021

Mayuri Thatikonda

**Document Control**

### Change Record:

|  |  |  |  |
| --- | --- | --- | --- |
| **Version Date Author Comments** | | | |
| 0.1 | 10 -SEP  2021 | Mayuri | Introduction & Architecture defined |
| 0.2 | 15 – Sep -  2021 | Mayuri | Architecture & Architecture Description appended and updated |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **Reviews:** |  |  |  |
| **Version Date Reviewer** | | | |
| 0.2 | 12-Sep-  2021 | Mayuri | Document Content , Version Control and Unit Test Cases to be added |

**Approval Status**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Review Date** | **Reviewed By** | **Approved By** | **Comments** |
|  |  |  |  |  |

# Contents

Abstract

1. [Introduction 4](#_bookmark0)

1.1 [Why this Low-Level design document? 4](#_bookmark1)

1.2 [Scope 4](#_bookmark2)

1.3 Constraints……………………………………………………………4

1.4 Risks………………………………………………………………5

1.5 Out of Scope………………………………………………………5

1. [Technical specifications 5](#_bookmark3)

2.1 Dataset Overview………………………………………………….5

2.2 Input Schema……………………………………………………..6

2.3 Prediction………………………………………………………..6

2.4 Logging………………………………………………………….6

2.5 Database………………………………………………………6

2.6 Deployment…………………………………………………6

1. [Technology Stack 7](#_bookmark4)
2. Model Training/Validation Workflow…………………………………….8

**Abstract**

Insurance premium prediction project will help you to know the amount that to be paid under given circumstances.

# Introduction

## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Deep EHR System. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The main objective of the project is to predict the amount that to be paid under given data.

The dataset contains 1339 records of information.

The dataset information:

* Contain a person’s age, sex, Bmi, children,smoker,region.
* The dataset includes all important information not including null values.
* Automate and streamline provider workflow

An dataset contains pert users information, such as:

* Age
* Sex
* Bmi
* Children
* Smoker
* Region

## Scope

This software system will be a Web application This system will be designed to predict the insurance amount, using previous dataset records available. This system is designed to predict the insurance amount from users information such as age, sex, Bmi, children, smoker, region.

## Constraints

We will only be selecting a few of the dataset records for training purposes.

## Risks

Document specific risks that have been identified or that should be considered.

## Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

# Technical specifications

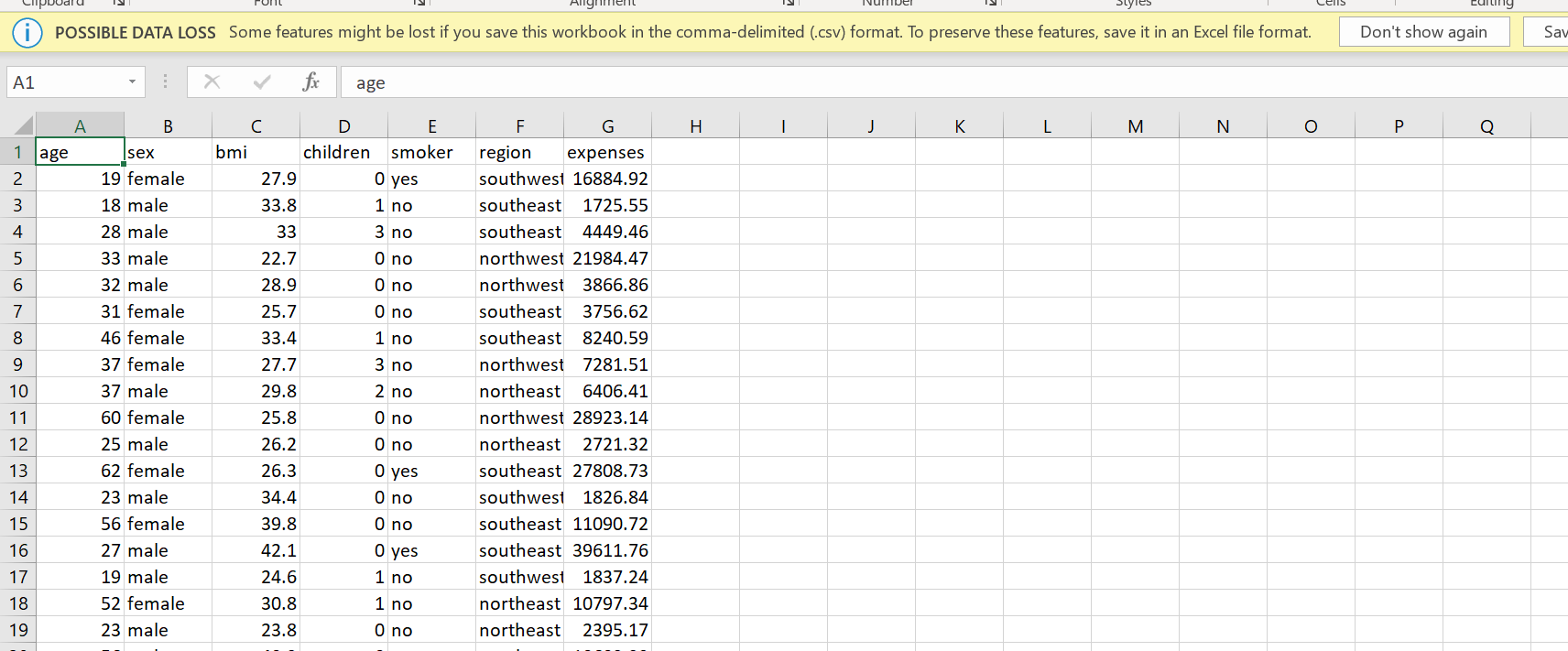
## 2.1. Dataset overview

Dataset consists of the user's personal information. There are total 1339 records are there.

There are a total of 1071 records in the training set and 267 records in the test set.

The data is in CSV format.

* Dataset table



## 2.2 Input schema

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature name** | **Datatype** | **Size** | **Null/Required** |
| Age | int | 3 | Required |
|  |  |  |  |
|  |  |  |  |

## 2.3 Predicting Disease

* The system displays the insurance amount.
* The User enters all the input fields.
* The User selects the region.
* The system presents the set of inputs required from the user.
* The user gives required information.
* The system should be able to predict insurance amount based on the user information.

## 2.4 Logging

We should be able to log every activity done by the user.

* The System identifies at what step logging required
* The System should be able to log each and every system flow.
* Developers can choose logging methods. You can choose database logging/ File logging as well.
* System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

## 2.5 Database

System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User chooses the disease.

2. The User gives required information.

3. The system stores each and every data given by the user or received on request to the database. Database you can choose your own choice whether MongoDB/ MySQL.

**2.6 Deployment**

1. Heroku



# Technology stack

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

# Model training/validation workflow

Data Cleaning

Data collection

Feature Engineering

Handling Categorical Variables

Feature Selection

Train test split

Model Training

Model Testing

Hyper Parameter Tuning

Model Deployment

Model Testing