High Level Design (HLD) Insurance Claim Probability

**Adil Shaikh**

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 27/02/2023 | 1.0 | Added Introduction and General  description | Adil Shaikh |
| 01/03/2023 | 1.1 | Added Design details and  abstract | Adil Shaikh |
| 12/03/2023 | 1.2 | Created and organized whole  document | Adil Shaikh |
| 15/03/2023 | 1.3 | Updated Design Details | Adil Shaikh |
| 21/03/2023 | 1.4 | Checked and added  Performance | Adil Shaikh |
| 22/03/2023 | 1.5 | Added Conclusion and  References | Adil Shaikh |
| 25/03/2023 | 2.0 | Updated diagrams in Design  Details | Adil Shaikh |

**Table of Content**

[Document Version Control 2](#_TOC_250015)

[Abstract 4](#_TOC_250014)

[Introduction 5](#_TOC_250013)

1. Why this High-Level Document? 5
2. [Scope 5](#_TOC_250012)
3. [Definition 5](#_TOC_250011)

[General Description 6](#_TOC_250010)

1. [Product Perspective 6](#_TOC_250009)
2. [Problem Statement 6](#_TOC_250008)
3. [Problem Solution 6](#_TOC_250007)
4. Proposed Methodology 6
5. Further Improvements 6
6. Data Required 6
7. Tools Used 7
8. Constraints 7
9. Assumptions 7

[Design Details 8](#_TOC_250006)

1. [Process Workflow 8](#_TOC_250005)
2. [Error Handling 9](#_TOC_250004)

[Performance 10](#_TOC_250003)

1. Reusability 10
2. [Application compatibility 10](#_TOC_250002)
3. Resources Utilization 10
4. [Deployment 10](#_TOC_250001)

[Conclusion 11](#_TOC_250000)

Reference 12

# Abstract

# More than two thirds of insurance businesses claim that predictive analytics has reduced their expenses and underwriting issues. With the advancement of artificial intelligence, insurance companies are increasingly turning to machine learning to achieve key objectives including cost reduction, enhanced underwriting, and fraud detection. In particular, insurance companies can use machine learning to quickly screen cases, fairly evaluate them, and calculate charges. This information emphasises how essential machine learning is for any insurance organisation. Insurance companies can predict their own profit and loss with ease by taking into account the likelihood of an insurance claim.

# Introduction

## Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) 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 the 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 features and the architecture of the project

## Scope

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

## Definition

This dataset has 7 features, as can be seen below:

1. Age: The age of the person claiming the insurance.
2. Gender: Gender of the policyholder (male = 1, female = 0).
3. BMI: By comparing weights to heights and finding weights that are too high or too low in relation to height, the body mass index (BMI) analyses body composition. The ideal range is between 18.5 and 25 steps. daily average of steps taken by the policyholder
4. Children: number of dependents or children of the policyholder
5. Smoker: Smoking status of the insured (non-smoker=0; smoker=1)
6. The primary residence of the policyholder in the United States is referred to as the region (northeast = 0, northwest = 1, southeast = 2, southwest = 3).
7. charges: Individual medical costs under health insurance

**General Description**

## **Product Perspective**

## Future projections hold a great deal of interest for insurance firms. The corporation has the opportunity to lessen its financial loss with accurate forecasting. Cost increases are primarily brought on by insurance companies' payment errors when processing claims. Reprocessing the claims makes up a large amount of administrative expenditures as a result of the payment problems.

## Problem Statement

## To develop a machine learning-based method for predicting insurance claims based on several factors.

## **Problem Solution**

## Create a web application to forecast insurance claims, which can assist citizens in determining whether or not a person can submit an insurance claim.

## **Data Required**

## I require the data, which consists of records of individuals with various parameters, in order to train the model.

## Data depends entirely on our problem definition.

## **Tools Used**

The entire model is built using the Python programming language and frameworks including NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn.

The IDEs used are PyCharm and Visual Studio Code.

The plots are shown using Matplotlib and Seaborn

Using Stream light, front end development is performed.

As a version control system, GitHub is employed.

I’ll be using AWS Amplify or S3 Bucket to host my ML model.

## **Constraints**

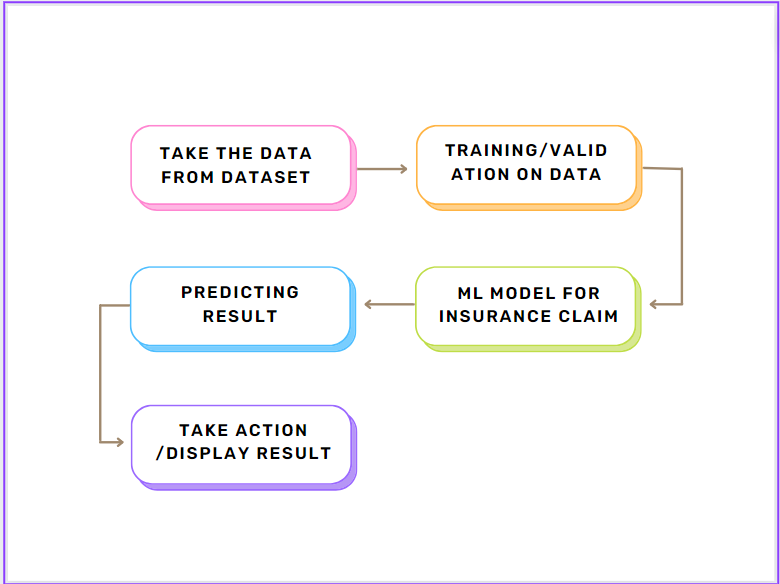
## The AQI prediction website ought to be simple to use. To develop various models for various Companies.

## **Assumptions**

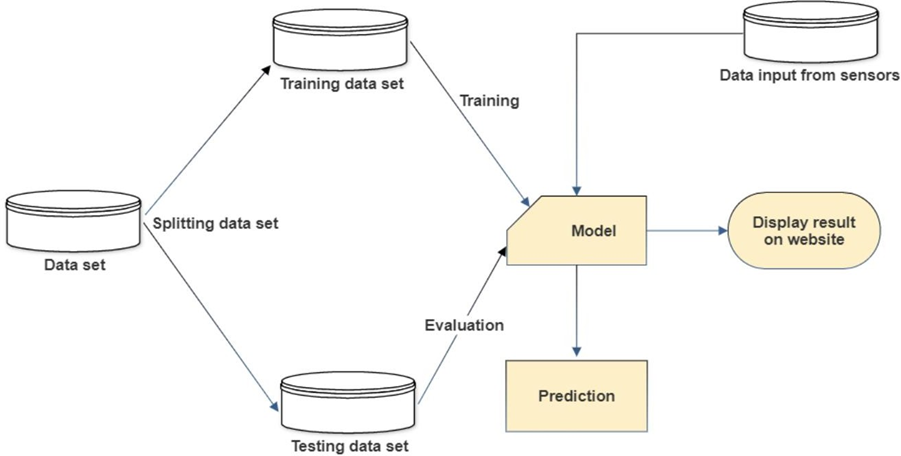
The project's major goal is to absorb real-time data from insurance providers and then estimate whether or not a person is qualified to make an insurance claim.

# Design Details

## **Process Workflow**

For identifying the different types of anomalies, we will use a machine learning model. Below is the process flow diagram.

Model Training and Evaluation



## **Error Handling**

I initially had a problem connecting to GitHub, which i fixed. The graphs were then not appearing correctly, and multiple texts were not being solved. The aforementioned mistakes have all been fixed.

I earlier I also have problem deploying my code on AWS because the charges are incurred but due the credits I earned by attending AWS seminar I managed to host the web app on Amplify.

I’ll be also try hosting the application on S3 bucket and the one which will be most efficient will be kept for a long time with least cost.

**Performance**

**Reusability**

Future projections hold a great deal of interest for insurance firms. The corporation has the opportunity to lessen its financial loss with accurate forecasting. Cost increases are primarily brought on by insurance companies' payment errors when processing claims. Reprocessing the claims makes up a large amount of administrative expenditures as a result of the payment problems.

**Application compatibility**

I adhere to application compatibility because I’m utilising Python, which is cross-platform compatible.

## **Deployment**

The code is available on Github and the readme file will assist you to run model on your local machine.

# Conclusion

The machine learning model for predicting insurance claims is proposed in this project. With the advancement of artificial intelligence, insurance companies are increasingly turning to machine learning to achieve key objectives including cost reduction, enhanced underwriting, and fraud detection. Also, it aids businesses in taking an insurance claim action.

Adil Shaikh