



High Level Design

Insurance Premium Prediction

Submission Date: 10/07/2024

Document Version Control

Date Issued	Version	Description	Author
30/03/2024	1.0	HLD 1.0	Md Rian Ur Rahaman

Contents:

Document Version Control Page 1

Contents..... Page 2

Abstract..... Page 3

Introduction..... Page 4

Genarel Description..... Page 5

Tools Used..... Page 6

Abstract:

This data science project focuses on developing a comprehensive solution to predict and mitigate customer churn in a subscription-based business model. The project follows a systematic end-to-end approach, encompassing data collection, preprocessing, exploratory data analysis (EDA), feature engineering, model selection, training, and deployment. The primary objective is to empower businesses with actionable insights to retain valuable customers and optimize revenue streams.

The project begins with data collection from diverse sources, including customer interactions, transaction history, and demographic information. Data preprocessing involves cleaning, handling missing values, and transforming raw data into a format suitable for analysis. The EDA phase employs statistical and visual techniques to uncover patterns, correlations, and potential factors influencing customer churn.

Feature engineering is a crucial step, extracting meaningful insights from the data to enhance predictive model performance. Various machine learning algorithms, such as logistic regression, decision trees, and ensemble methods, are explored and evaluated for their ability to predict customer churn accurately. Model hyperparameter tuning and cross-validation techniques are applied to optimize the model's performance.

The project emphasizes the importance of interpretability, providing stakeholders with clear insights into the factors contributing to customer churn. Model explanations and visualizations aid in understanding the decision-making process and guide strategic decision-making.

The selected model is then deployed into a production environment, allowing real-time predictions on new data. Integration with existing systems and continuous monitoring ensure the model's effectiveness over time.

The project concludes with a comprehensive evaluation of the entire process, including model performance metrics, business impact, and insights gained. By implementing this end-to-end data science project, businesses can proactively address customer churn, reduce revenue loss, and enhance overall customer satisfaction.

INTRODUCTION

1.1 Why this Level Design Document?

The purpose of this High-Level 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 the design aspects and define them in detail
- . Describe the user interface being implemented
- . List and describe the non-functional attributes like:
 - . Security
 - . Reliability
 - . Maintainability
 - . Portability
 - . Reusability
 - . Application Compatibility
 - . Resource Utilization

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.

a. 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.

b. Product Perspective:

I create an end-to-end project solution and web app which able to predict the premium of the personal for health insurance.

c. Approach:

The classical machine learning tasks like data exploration, data cleaning, feature engineering, model building and model testing. Trying different machine learning algorithms that's best fit for the above case.

Some Famous Algorithms like linear regression, Decision Tre Regression and Gradient Boosting, XG-boost Regression were used in this project.

d. Create API or User Interface:

I created a user interface using HTML, CSS and Flask.

e. Logging:

Logging is created in the project which is very useful to detect errors or result.

f. ML-Ops Pipeline:

Successfully use ops pipeline for delivery like: CI/CD pipeline, DVC, ML-flow, in this project.

Page 5

g. Deployment:

I host my model in the AWS cloud. **Unfortunately, they charged me many rupees so I delete my ECR repository and AWS app runner.**

h. Tools Used:

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



