High Level Design (HLD)

Insurance Premium Prediction

Document Version Control

Date Issued	Version	Description	Author
9/6/23	1.0	Initial HLD	Dilip Kumar Panda

Contents

Abstract

Health insurance premium prediction is a vital task in the insurance industry, as accurate estimation of premiums helps both insurers and policyholders make informed decisions. This study focuses on developing a predictive model to estimate health insurance premiums based on various factors such as age, gender, BMI, smoking status, number of children and geographical location.

The dataset used for this research comprises a large sample of historical health insurance data, encompassing diverse demographic and health-related attributes. Machine learning algorithms, including linear regression, decision trees, and ensemble methods, were employed to analyze the dataset and build the predictive model.

Feature engineering techniques were applied to handle missing data, normalize variables, and create meaningful new features. The model was trained and evaluated using appropriate performance metrics, such as mean absolute error and root mean squared error, to assess its accuracy and generalization capabilities.

The results demonstrate that the developed predictive model performs well in estimating health insurance premiums. The model achieves a high degree of accuracy, capturing the complex relationships between predictors and premiums. Moreover, it provides valuable insights into the relative importance of various factors in determining insurance costs.

This research contributes to the field of health insurance by providing a robust predictive model that aids insurance providers in accurately estimating premiums. The model's effectiveness empowers individuals to make informed decisions regarding their health insurance coverage and allows insurers to effectively price their policies, leading to fairer and more efficient insurance markets.

Introduction

The Health Insurance Premium Prediction project aims to develop a predictive model that accurately estimates health insurance premiums based on various factors. This High-Level Design Document (HLDD) provides an overview of the system's architectural design and components, outlining the structure and interactions necessary to achieve the desired prediction functionality.

The primary objective of this document is to present a conceptual understanding of the system's design to developers, designers, and stakeholders involved in the project. By outlining the high-level architecture and design principles, this document serves as a foundation for effective collaboration, decision-making, and implementation planning throughout the software development lifecycle.

In this project, we recognize the importance of accurate premium estimation in the health insurance industry. The ability to predict premiums based on factors such as age, gender, BMI, smoking status, and geographical location empowers both insurers and policyholders to make informed decisions regarding insurance coverage.

The HLDD will outline the system architecture, data flow, interfaces, functional components, algorithms, security considerations, performance considerations, error handling, deployment strategy, and any dependencies or constraints. By addressing these key aspects, we aim to create a comprehensive and robust system that provides accurate premium predictions.

It is crucial to note that this document focuses on the high-level design of the system and does not delve into specific implementation details. For developers seeking more granular information, detailed design documents, code documentation, and relevant resources will provide the necessary guidance.

By following the architectural principles and guidelines outlined in this HLDD, we strive to develop a powerful and reliable health insurance premium prediction system. This system will not only assist insurance providers in accurately estimating premiums, but it will also empower individuals to make informed decisions about their health insurance coverage.

With this HLDD as our guide, we commit ourselves to designing and implementing a solution that meets the needs of our stakeholders, improves decision-making, and enhances the overall health insurance experience.

General Description

Problem statement :

The problem addressed by the Health Insurance Premium Prediction project is the lack of accurate and transparent methods for estimating health insurance premiums. Current approaches rely on generalized rates or manual calculations, leading to inconsistencies and inaccuracies in pricing. This results in policyholders overpaying or insurance providers facing financial losses. There is a need for a data-driven predictive model that considers relevant factors and provides precise premium estimations. This project aims to develop a solution that improves fairness, accuracy, and efficiency in the health insurance industry, empowering stakeholders to make informed decisions about their coverage.

Proposed Solution:

The proposed solution is to develop a predictive model using machine learning techniques to accurately estimate health insurance premiums. By analyzing factors such as age, gender, BMI, smoking status, and geographical location, the model will provide transparent and data-driven premium calculations. This solution improves fairness, accuracy, and decision-making for both policyholders and insurance providers in the health insurance industry.

Improvement:

To enhance the predictive model for health insurance premium estimation, we can incorporate additional relevant features, fine-tune the algorithms, address data imbalances, explore nonlinear relationships, apply regularization and feature selection techniques, implement continuous model monitoring and updating, and integrate external data sources. These improvements will lead to more accurate and reliable premium predictions, benefiting both policyholders and insurance providers.

Data Requirements:

I have taken the dataset from Kaggle which contains around 1300 rows and 7 columns.

Tools Used:

Vs-Code is used as IDE

For visualization matplotlib as seaborn is used

For deployment Azure is used

Github is used as version control system

Jupyter notebook is used as initial model evaluation

Html is used for front end web app

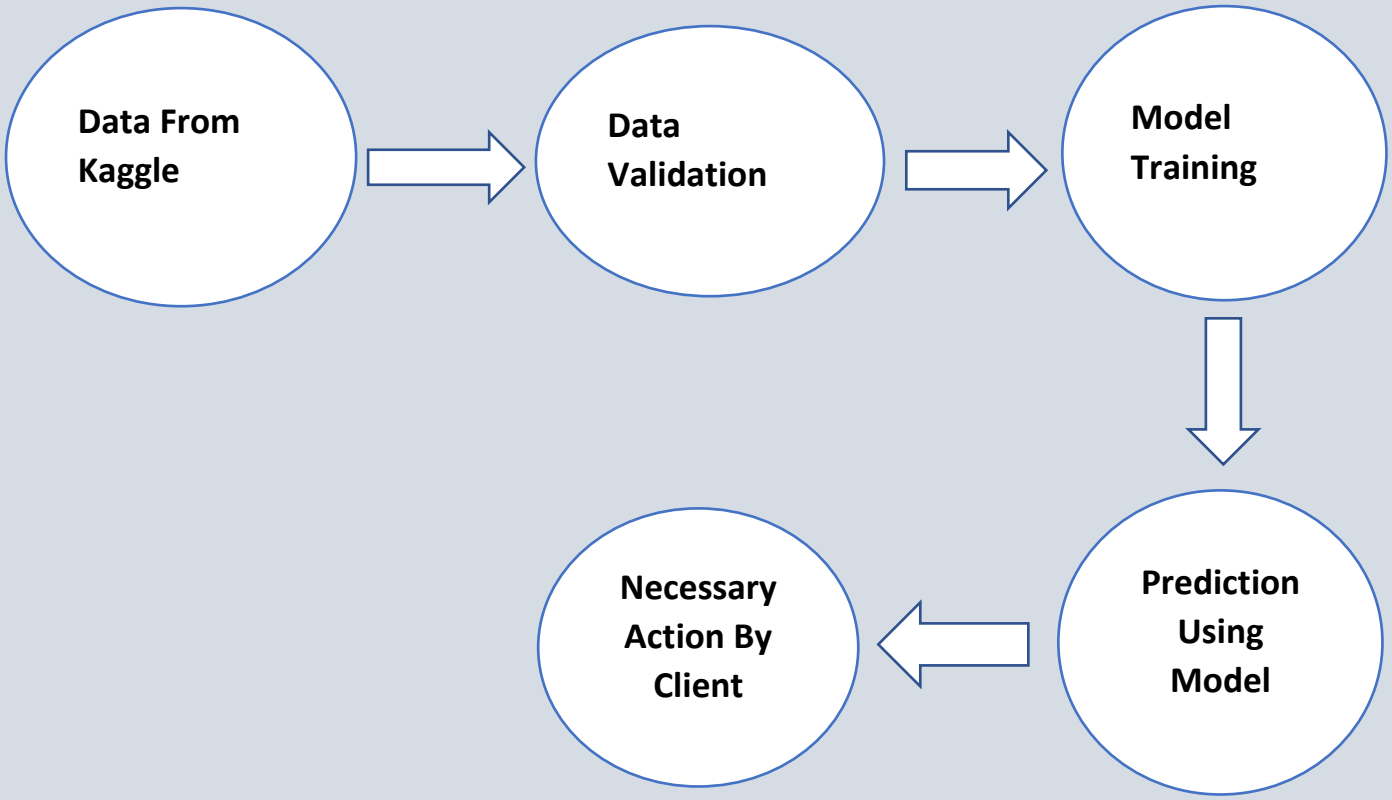
Flask is used to create a web app

Assumption:

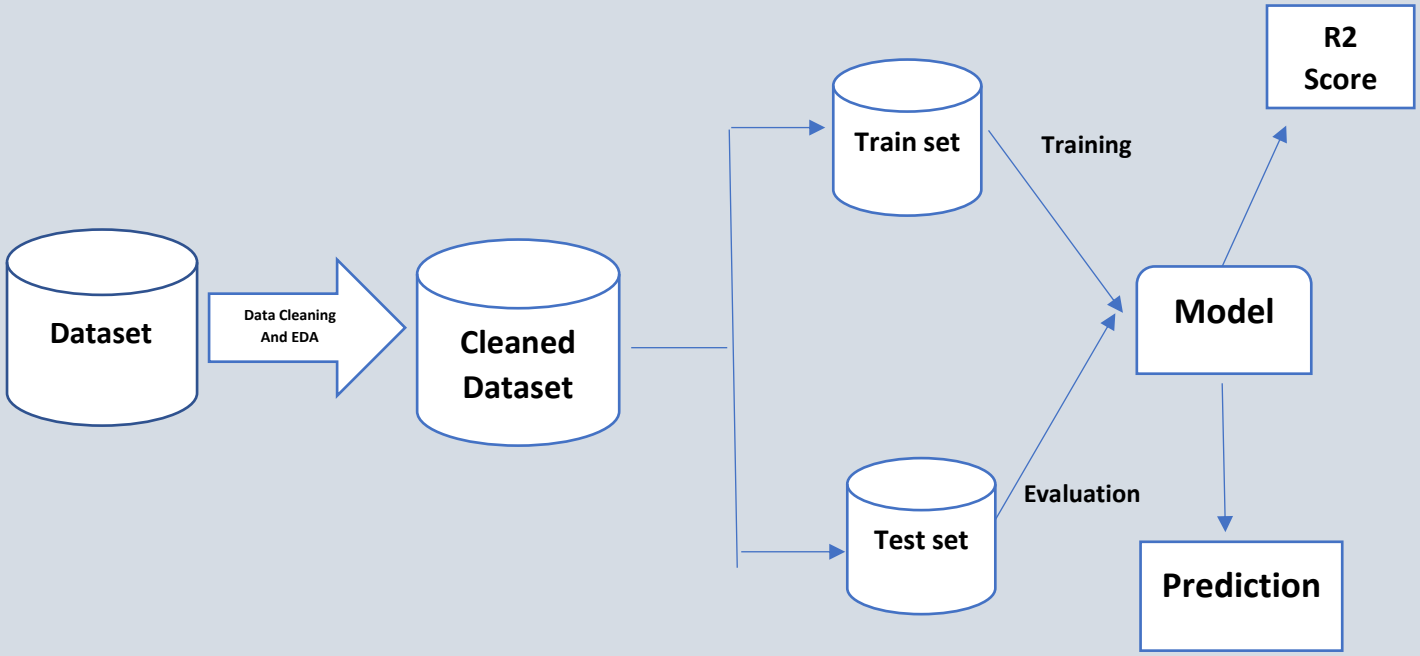
The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through Insurance premium Prediction. Machine Learning based model is used for Predicting the above-mentioned use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

Design Detail

Proposed Methodology:



Model Training And Evaluation :



Deployment Process:

Event Log:

The project is created in such a way that all the logs can be stored in a log folder which can be accessed whenever required.

The logs are maintained as per date and time which is easy to find any cache in the model during execution

Logging is done in every step of the project while building the model.

Error handling:

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage. We have designed for program in such a way that it will display in which line and in which Python script the error has been occurred.

Application Compatibility:

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

Reusability:

The Code is written in a way that it can be reused or reproduced without any complication in any operating system.

Performance:

The model Accuracy is about 88 percent in the train data set and about 86 percent in the test dataset which suggest us that the model performs well in test case scenario.

Deployment:

For deployment We have used Microsoft azure using Github actions.

Conclusion

In conclusion, this High-Level Design Document (HLDD) has outlined the architectural design and components of the Health Insurance Premium Prediction system. The proposed solution leverages machine learning techniques to provide accurate and transparent premium predictions based on various factors. By addressing the challenges of inaccurate premium estimations in the health insurance industry, this system aims to improve fairness, accuracy, and decision-making for policyholders and insurance providers.

The document has also highlighted potential improvements such as incorporating additional features, fine-tuning algorithms, addressing data imbalances, and implementing continuous model monitoring. Adhering to constraints such as data availability, privacy, and regulatory compliance ensures the system's ethical and practical applicability. With this HLDD as a guide, the project can confidently proceed with the development and implementation of a robust Health Insurance Premium Prediction system, ultimately enhancing the insurance experience for all stakeholders involved.

References

Wikipedia

Reference HLDD

ChatGPT

Google images