High-Level Design (HLD)

Insurance Premium Prediction

Revision Number – 1.0

Last Date of Revision: 14 – 05 -2023

Santhana Lakshmi

Document Version Control

Date	Version	Description	Author
14– 05 - 2023	1.0	First Draft	Santhana Lakshmi

Contents

Abstract

This High-Level Design (HLD) document outlines the design of an Insurance Premium Prediction system. The system aims to predict the insurance premium for customers using machine learning algorithms, based on their personal and vehicle information. The document describes the proposed solution, technical and data requirements, tools and technologies used, design details, and deployment process.

1.Introduction

The Insurance Premium Prediction system is designed to provide accurate insurance premium quotes for customers based on their personal and vehicle information. The traditional method of manually calculating premiums can be time-consuming and prone to errors. Therefore, the proposed solution is to use machine learning algorithms to analyze historical data and generate accurate quotes in real-time.

1.1 Why this High-Level Design Document?

This document serves as a blueprint for the development team to understand the proposed solution and how the system will be built.

1.2 Scope

The system will be designed to predict insurance premiums for customers based on their personal and vehicle information. The system will use machine learning algorithms to analyze historical data and generate accurate quotes.

1.3 Definitions

Insurance premium: The amount of money paid by a customer to an insurance company for a policy.

Machine learning: A subset of artificial intelligence that uses algorithms to learn from data and make predictions or decisions.

2. General Description

The Insurance Premium Prediction system is designed to predict insurance premiums for customers based on their personal and vehicle information. The system uses machine learning algorithms to analyze historical data and generate accurate quotes. The system is a standalone application that integrates with the insurance company's existing systems to generate insurance premium quotes for customers. The system follows a process flow of data collection and pre-processing, feature engineering, model training and testing, and model deployment. The system is designed to handle large amounts of data and generate insurance premium quotes in real-time. The system is also designed to be reusable and scalable to handle different types of insurance policies and customer information. The user interface is user-friendly for inputting customer and vehicle information and generating insurance premium quotes. The system will be deployed on the cloud to allow for scalability and accessibility. Overall, the Insurance Premium Prediction system aims to address the challenge of generating accurate insurance premium quotes for customers in a timely and efficient manner.

2.1 Product perspective

The system will be a standalone application that integrates with the insurance company's existing systems to generate insurance premium quotes for customers.

2.2 Problem statement

The insurance industry faces the challenge of generating accurate insurance premium quotes for customers based on their personal and vehicle information. Traditional methods of manually calculating premiums can be time-consuming and prone to errors.

2.3 Proposed solution

The proposed solution is to use machine learning algorithms to analyze historical data and generate accurate quotes for customers.

2.4 Technical requirements

The technical requirements for the Insurance Premium Prediction system include:

Programming Language: Python will be used as the primary programming language for building the system.

Machine Learning Libraries: Scikit-learn, Pandas, NumPy, and Matplotlib will be used as the primary machine learning libraries to develop and train the predictive models.

Jupyter Notebook: Jupyter Notebook will be used for data exploration, feature engineering, and model training.

Data Storage: The system will require a database to store customer information and historical policy data.

Cloud Computing Platform: A cloud computing platform like Streamlit will be used to deploy the system and to ensure scalability and availability.

APIs: Application Programming Interfaces (APIs) will be used to integrate the system with the insurance company's existing systems and to receive data input from the user.

2.5 Data requirements

The Insurance Premium Prediction system will require the following data for generating accurate quotes:

Customer personal information(age, gender, etc.): The system will require information such as age, gender, occupation, marital status, and driving history of the customer.

Vehicle information (make, model, year, etc.): The system will require information such as the make, model, year, engine size, and safety features of the vehicle.

Historical insurance policy data: The system will require data on past insurance policies, including the type of coverage, premium amounts, and claims history.

2.6 Tools and technologies used

Python: Python is a high-level programming language that is widely used for machine learning and data analysis.

Scikit-learn: Scikit-learn is a machine learning library in Python that provides simple and efficient tools for data mining and data analysis.

Pandas: Pandas is a Python library for data manipulation and analysis.

NumPy: NumPy is a Python library for scientific computing that provides support for large, multi-dimensional arrays and matrices.

Matplotlib: Matplotlib is a Python library for data visualization and plotting.

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows users to create and share documents that contain live code, equations, visualizations, and narrative text.

2.7 Constraints

The system must be scalable to handle large amounts of data and capable of generating insurance premium quotes in real-time. The system must maintain the confidentiality, integrity, and availability of customer data. The system should have security measures such as access controls, encryption, and data backups. The system should be compatible with the insurance company's existing systems and databases. It should be able to integrate seamlessly with existing systems without disrupting the workflow.

2.8 Assumptions

Sufficient historical data is available for training and testing the machine learning algorithms. The model will be trained on the assumption that historical data accurately represents current and future trends. The system will be integrated with the insurance company's existing systems for seamless deployment. The system will be able to handle large amounts of data in real-time. The system will be deployed on the cloud for scalability and accessibility. The machine learning algorithms used in the system will be effective in accurately predicting insurance premiums. The system will be able to handle various types of insurance policies and customer information. The performance of the system will depend on the quality and quantity of input data. The system will adhere to industry standards and regulations regarding the use of customer data and privacy.

3.0 Design Details

3.1 Process flow

Data collection and pre-processing: The system will collect customer and vehicle information from various sources such as online forms, databases, and insurance policies. The data will be pre-processed to remove any missing values or errors.

Feature engineering: The pre-processed data will be used to create new features such as age of the vehicle, driving history, and previous insurance claims. These features will be used as input variables for the machine learning algorithms.

Model training and testing: The system will use machine learning algorithms such as linear regression, decision trees, and neural networks to train and test the models. The models will be evaluated based on metrics such as accuracy and mean absolute error.

Model deployment: The best-performing model will be deployed as a standalone application and integrated with the insurance company's existing systems to generate insurance premium quotes.

3.2 Event log

The system will log events such as data collection, feature engineering, model training, and model deployment for monitoring and troubleshooting purposes.

3.3 Error handling

The system will handle errors such as missing data, invalid data, and model prediction errors. Error messages will be logged and displayed to the user.

3.4 Performance

The system will be designed to handle large amounts of data and generate insurance premium quotes in real-time.

3.5 Reusability

The system will be designed to be reusable and scalable to handle different types of insurance policies and customer information.

3.6 Application compatibility

The system will be compatible with the insurance company's existing systems and databases.

3.7 Recourse utilization

The system will be designed to use minimal resources such as CPU and memory.

3.8 Deployment

The system will be deployed on the cloud to allow for scalability and accessibility.

4. Dashboards:

A dashboard is a visual representation of data that displays key metrics and information in an easily understandable format. It is typically designed to provide a quick and easy overview of the performance of a business or specific department. Dashboards can be customized to show relevant data and KPIs, such as sales, revenue, customer satisfaction, and productivity.

4.1 Key Performance Indicators (KPIs)

Key Performance Indicators (KPIs) for Insurance Premium Prediction can vary depending on the specific goals and objectives of the insurance company.

5. Conclusion

This HLD document outlines the design of an Insurance Premium Prediction system that uses machine learning algorithms to generate accurate insurance premium quotes for customers based on their personal and vehicle information. The system will be designed to be scalable, reusable, and user-friendly.