Architecture

Insurance Premium Prediction

Revision Number – 1.0

Last Date of Revision – 01-09-2023

Anuj Dhyani

Document Version Control

Date	Version	Description	Author
02-09-2023	1.0	Abstract	Anuj Dhyani
02-09-2023	1.1	Architecture	Anuj Dhyani
03-09-2023	1.2	Architecture Design	Anuj Dhyani

Contents

Abstract

Our primary objective is to empower individuals to approach health insurance providers with confidence, armed with a clear understanding of the expected financial commitments tied to their selected coverage. By eliminating the uncertainty surrounding insurance premiums, we enable people to prioritize their health and well-being over pricing complexities.

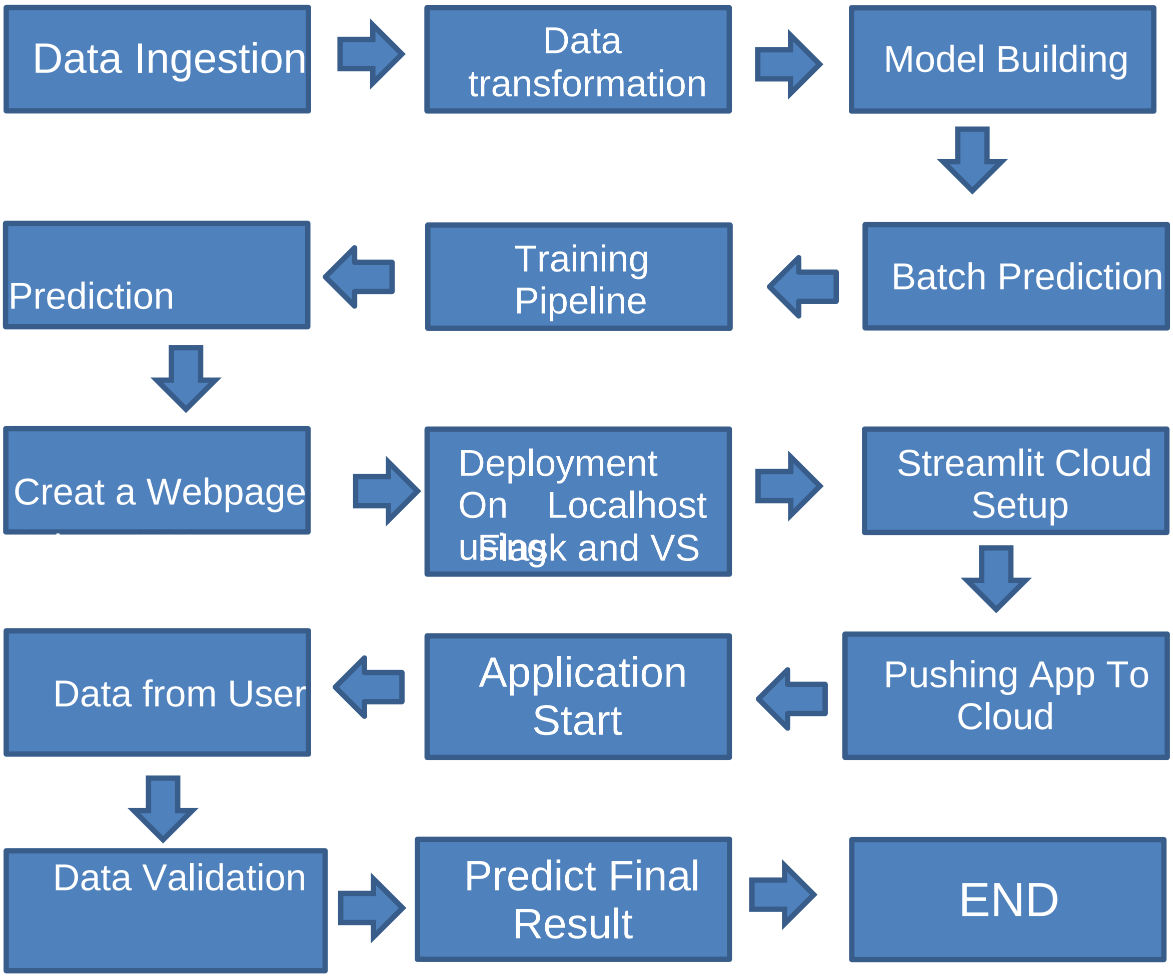
This innovative approach shifts the focus away from the intricacies of policy pricing and toward a user-centric perspective that promotes informed decision-making. Ultimately, our goal is to enhance the overall insurance experience, making it more personalized and user-friendly. We aim to facilitate insurance choices that reflect individual needs and financial considerations, empowering individuals to invest in their health with clarity and assurance.

Introduction

Why this Architecture Design documentation?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the Insurance Premium Prediction code. The Architecture design documentation is designed in such a way that the programmer can directly code after reading each module description in the documentation.

1 Architecture



2 Architecture design

The architecture for the Insurance Premium Prediction project comprises data collection and preprocessing from various sources, storage in a database system, data analysis and feature selection using machine learning frameworks, model development with algorithm selection and training, real-time data integration for up-to-date predictions, and model deployment through containerization and microservices. A user-friendly web interface will enable users to input product details and receive accurate price predictions, facilitating informed shopping decisions. This comprehensive architecture ensures the project's ability to provide real-time and reliable price predictions for store products, enhancing the overall shopping experience.

3 Data gathering from main source

The data for the current project is being gathered from the Kaggle dataset, which is available at the following link: Insurance Premium Prediction | Kaggle. This dataset serves as the primary source for our project's data analysis and premium prediction tasks.

3.1 Data description

We have train (1070) and test (268) data set, train data set has both input and output

Columns Are :

1. **Age**: Age of the insured individuals, a key factor in premium calculation due to its influence on health risk.

2. **Sex**: Gender of the insured individuals, impacting premiums based on gender-specific health risks.

3. **BMI (Body Mass Index)**: Measure of body weight relative to height, influencing premiums based on health implications.

4. **Children**: Number of dependents covered, affecting policy costs due to family size.

5. **Smoker**: Smoking status (yes/no), a significant factor in premium pricing due to health risks associated with smoking.

6. **Region**: Geographic location of the insured, which can affect healthcare costs and insurance pricing.

7. **Expenses**: Actual medical expenses incurred, providing insight into healthcare costs and utilization for premium calculation.

3.2 Data Ingestion

The cornerstone of our data-driven project was established through a systematic process of data acquisition and ingestion. Utilizing Kaggle, a reputable platform renowned for its high-quality datasets, we identified and acquired the crucial data required for our price prediction project. This dataset, integral to our goal of accurate price forecasting, was meticulously downloaded and securely stored within our local system infrastructure. Subsequently, we initiated the data ingestion phase, where the dataset seamlessly integrated into our project's data pipeline. This meticulous approach ensures that our project is built upon a solid foundation, setting the stage for robust and precise price prediction models and analysis.

3.3 Data pre-processing

Steps performed in pre-processing are:

- First read data from Artifact folder
- Checking unnecessary columns
- One column has product id which is unique for every product so I deleted that column.
- Checked for null values
- there are too many null values are present in two columns that's why I deleted them
- Performed one-hot encoder on categorical columns.
- Perform Ordinal Encoder on Ordinal Columns.
- Scaling is performed for needed information.
- And, the info is prepared for passing to the machine learning formula

3.4 Modelling

The pre-processed information is then envisioned and every one the specified insights are being drawn. though from the drawn insights, the info is at randomunfold however still modelling is performed with completely different machinelearning algorithms to form positive we tend to cowl all the chances. and eventually, Gradient Boosting performed well

3.5 UI integration

Both CSS and HTML files are being created and are being integrated with the created machine learning model. All the required files are then integrated to theapp.py file and tested locally

3.6 Data from user

The data from the user is retrieved from the created HTML web page and Streamlit webpage.

.

3.7 Data validation

The data provided by the user is then being processed by app.py and application.py file and validated. The validated data is then sent for the prediction.

.

3.8 Rendering the results

The data sent for the prediction is then rendered to the web page.

3.9 Deployment

The tested model is then deployed Streamlit Cloud and local machine. So, users can access the project from any internet devices.