INSURANCE PREMIUM PREDICTION WITH AVAILABLE RISK FACTORS

Presented By:

Amit Pingale

Himani Solanki

Shubham Patel

Overview:

The project will take the dataset consisting of various independent variables and using various regression models we will produce computational matrix. The best regression model will be selected and will be used for determining whether the insurance will be categorized into silver, gold, and platinum insurance coverages.

The insurance company will upload the data using a web application. This will help in effective and remote communication for serving clients located at different places.

Goal:

Accurately predict insurance premium costs for a person knowing which factors(region, smoking/drinking habits, age, body mass index) are having importance and weight.

Use Cases:

Insurance agents could use our web application to insert data and have a prediction of insurance premium of a particular individual by having their location, smoking/drinking habits, age,etc.

Data:

Columns - age: age of primary beneficiary

- **sex**: insurance contractor gender, female, male
- **bmi**: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
- **children**: Number of children covered by health insurance / Number of dependents
- **smoker**: Smoking
- **region**: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
- **charges**: Individual medical costs billed by health insurance

https://www.kaggle.com/mirichoi0218/insurance/data

Process Outline:
- Preprocessing
    - Data Cleaning

- Handling Missing values
- Exploring Data Analysis
- Study of Supervised approaches and select the best model for prediction
- Design of a pipeline and system to implement this approach and discussion on the system's capabilities
- Deploy the model on AWS
- Build a web application to demonstrate the prediction and recommendation results
- Dockerise the pipeline

Milestones:

Timeframe	Delivery
Day 1-2	Data Preprocessing and Exploratory Data Analysis
Day 3-6	Model Building, Training, Selection
Day 7-8	Deployment of models on cloud and build web application
Day 9-10	System integration and documentation

Personnas:

The Insurance companies could predict insurance of a particular individual person. The premium would depend on various factors like region, smoking/drinking habits, age, body mass index,ets

Deployment details:

- **Languages:** Python
- **Pipeline:** Luigi/skLearn
- **Container:** Docker
- **Cloud Tool:** AWS(Amazon Web Service) EC2/S3

References:

https://elitedatascience.com/machine-learning-projects-for-beginners#health-care