**INSURANCE CLAIMS PREDICTION**

**Objective:**

The primary objective of this study is to utilize data analytics techniques to enhance business decision-making processes. The study aims to identify patterns and insights from **Predict Insurance claim Dataset** to support strategic planning, optimize operations, and improve in detection of fraud and access risks and improve customer service.

**Rationale:**

In today’s era is to enhance the accuracy and efficiency of risk management, fraud detection and claim processing. Data analytics offers powerful tools and methodologies to extract valuable information, enabling the performance and fine-tune it for accuracy. This study aims to demonstrate the **Prediction of a outcomes in insurance claim data, gather and clean data and then select relevant features.**

**Methodology:**

**1. Data Collection and Preprocessing:**

Data collection and its pre-processing plays a crucial role in data analytics and it involves certain steps which are as follows Data Cleaning: The data was almost perfect to begin with analyzation.

**Handling Missing Values:**

* Identify and analyse columns with missing values.
* Decide on appropriate strategies for handling missing data (e.g., imputation using mean, median, mode; deletion of rows/columns; predictive filling based on other features).

**Handling Duplicates:**

* Identify and remove duplicate records from the dataset if necessary.
* Ensure uniqueness in identifiers (e.g., Ages) to avoid data redundancy.

**Data Integration:**

* The data was already found in an integrated csv file.

**2. Data Analysis:**

Data Analysis provides valuable insights into the characteristics and relationships within the dataset, laying a solid foundation for developing predictive models in a customer purchase prediction project. By thoroughly understanding the data through EDA, data analysts and scientists can make informed decisions about feature selection, data preprocessing, and model development, ultimately enhancing the accuracy and effectiveness of predicting Insurance claim

For my data analysis I have imported different python libraries like Pandas, NumPy, Seaborn Matplotlib and Sklearn. I have used pandas for importing my dataset in my Jupyter notebook. Then I described the data for checking its shape and I used matplotlib for my visualization of variables.

**3. Model Development:**

**Algorithm Selection:-**

Linear Regression, Decision Tree Classifier, Random Forest, Support vector machine (SVM), F1 Score

**Training and Validation: -**

Split the dataset into training and validation sets. Train models on the training set and evaluate their performance on the validation set.

Following were the train and test set splits:

X\_train shape: (1069, 7)

X\_test shape: (268,7)

Y\_train shape: (1069,)

Y\_test shape: (268,)

Note: “May vary during the run of code.”

**Model Tuning:** -

Optimize model parameters to enhance accuracy and robustness.

**4. Implementation and Testing:**

**Pilot Study:** A set of data was entered in the model after the evaluation of the algorithms and the result was fairly acceptable.

**Dataset:**

This dataset contains comprehensive information on insurance data and behaviour, it contains policy details, claim amounts and dates. It includes customer demographics, This data is used to analyze and predict risks, detect fraud and improve claims processing efficiency.

**Features:**

* **Age**: Number of ages of customeror policy holder or a individual filing the claim.
* **Bmi:** Body Mass Index (BMI) used to access weight and height to evaluate health risks
* **Region:** Refers to geographical area where policy holder resides or where claim occurred
* **Charges:** Refers to the costs associated with medical treatments, repair or services covered under policy
* **Insurance Claim:** It includes records of incidents where the policy holder request compensation from their insurance provider
* **Smoker:** the individuals identified as tobacco users or cigarette users
* **Children’s:** Dependents covered under a policy such as minors who are beneficiaries of health or life insurance.
* **Sex:** Gender of policy holders and claimant such as male, female, others

**Target:**

Target Variable is predicting claim amounts and identifying fraudulent claims

**Expected Outcomes:**

* It gives accurately predicting claim amounts and efficiently detecting and preventing fraud.
* Improving minimizing of financial losses.
* It enhances customer Satisfaction through prompt and fair settlements

**Conclusion:**

This study aims to showcase the tool, techniques and methodologies of Data Analytics in upgrading the quality and visibility of the predict the insurance claim and risk factors patterns of claims behaviour and effectiveness of fraud detection measures enhance the customer satisfaction and optimize their business processes.