**Title:**

**Predictive Modeling to Expand Vehicle Insurance Customer Base**

**Problem Statement:**

Our client, an insurance company, is looking to broaden its product offerings by predicting which of its existing health insurance policyholders might be interested in purchasing vehicle insurance. They seek to optimize their communication strategy, engage customers more effectively, and boost revenue.

**Business Objective:**

1. Identifying Potential Customers: We aim to create a predictive model to identify current health insurance policyholders who are likely to show interest in buying vehicle insurance.
2. Effective Communication: By pinpointing potential customers, we can tailor our marketing and communication efforts to maximize conversion rates.

**Predictive Modeling Task (Classification):**

**Data Mining:**

Key Components of Data Mining

* balancing the target variable,
* encoding categorical variables
* normalizing the data
* Assessing the quality and performance of models using metrics like accuracy, precision, recall and F1 score

We will also perform exploratory data analysis (EDA) to gain insights

we will gain insights by drawing the following graphs:

* Bar charts
* Pie charts
* Statistic of the data
* Comparison Graph

**Methodology**

Our core task involves predictive modeling, specifically a binary classification problem. We will classify existing health insurance policyholders into two groups:

* Those "Interested in Vehicle Insurance."
* Those "Not Interested in Vehicle Insurance."

A predictive modeling classification task involves building a model that will classify data into predefined categories or classes based on input features. The goal is to use historical data, where the classes are known, to train a model that can accurately predict the class labels for new, unseen data.

**Dataset Size:**

Dataset contains more than 3 lacs observations so will split the dataset and we will shrink dataset by applying different data mining techniques the dataset will also be shrink when we remove missing values and outliers from the dataset

**Training and Validation:**

We'll allocate a significant portion of the dataset, approximately 70-80%, for model training. The remaining 20-30% will serve for validation and testing to evaluate model performance.

**Predictor Variables:**

Predictor Variables include:

* Gender
* Age
* Driving License
* Region Code Type
* Previously Insured
* Vehicle Age
* Damage Status
* Premium Amount
* Sourcing Channel
* Vintage

**Target Variable:**

Our binary class target variable will be "Response":

* 1 signifies customers interested in vehicle insurance.
* 0 represents customers uninterested in vehicle insurance.

**Data Mining Techniques:**

We will employ various data analysis methods suitable for binary classification tasks:

* Decision Trees
* Logistic Regression
* Neural Networks (if deemed suitable)

**Tools:**

Our primary tool will be Python, supplemented by libraries such as scikit-learn, pandas, for data preparation, model development, and evaluation. And we will use KNIME tool for visualizing patterns and insights to make them understandable to stakeholders.

**Dataset URL:**

[**https://www.kaggle.com/code/yashvi/vehicle-insurance-eda-and-boosting-models/input?select=test.csv**](https://www.kaggle.com/code/yashvi/vehicle-insurance-eda-and-boosting-models/input?select=test.csv)

**Team Members:**

This project will be led by:

- [Your Name]

- [Team Member 2 Name (if applicable)]

Our goal is to harness predictive modeling to enhance our client's revenue by identifying potential vehicle insurance customers. We are dedicated to delivering valuable insights and an effective predictive model to accomplish these objectives.