Title:

Health Insurance Cross Sell Prediction

Problem Statement:

Our client, an insurance company, is interested in identifying potential customers who may be interested in their vehicle insurance product. Specifically, they want to predict whether customers who have purchased health insurance from them in the past would also be interested in purchasing vehicle insurance from them. The client has provided us with a dataset containing demographic information, vehicle-related information, and policy-related information for their customers, along with a binary indicator for whether each customer is interested in vehicle insurance or not. Our task is to build a model that can accurately predict this binary indicator for new customers.

Approach:

The project began by importing the necessary Python libraries and loading the training and test data. The data was then preprocessed by checking for null values and encoding categorical variables using OneHotEncoder. The target variable was then extracted, and the remaining data was split into training and validation sets using the train_test_split function.

Logistic Regression was used as the predictive modeling algorithm. The model was trained on the training data and evaluated on the validation data using accuracy as the evaluation metric. The trained model was then used to predict the target variable for the test data. The predictions were then saved in a submission file in the required format.

The accuracy of the logistic regression model on the validation set was found to be 0.8765, indicating good performance. The submission file showed that out of 127,037 predictions, 126,994 were predicted as 0 (not interested in vehicle insurance), and only 43 were predicted as 1 (interested in vehicle insurance).

Conclusion:

In this project, a logistic regression model was used to predict the interest of health insurance policyholders in vehicle insurance policy. The model showed good performance on the validation set and accurately predicted the target variable for the test data. The submission file showed that very few policyholders were predicted to be interested in the vehicle insurance policy. This may indicate that the company needs to re-evaluate its vehicle insurance policy to make it more attractive to customers.