Capstone Project Submission

Team Member's Name Contribution and Email:

Mohd Sharik - livisliquoro999@gmail.com

1. Code Notebook :
    a. EDA : In-depth Analysis of each Feature
    b. Outliers handling
    c. Data preprocessing
    d. Model implementation
    e. Model explainability
    f. Note book summaries.
    ,
2. Documentation
    a. Presentation
    b. Technical Documentation
    c. Project summary
    d. Readme file

Google Drive Link: -

https://drive.google.com/file/d/1eWEYexYpvnVN31yR3COy96IkXCtNfYtS/view?usp=sharing

GitHub Link: -

https://github.com/MohdSharik99/Capstone-Project-HEALTH-INSURANCE-CROSS-SELL-PREDICTION---Classification

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions.

Problem Statement:
The task is to use existing Health and Vehicle Insurance Customer Data to predict whether the any new Customers are open to purchasing Vehicle Insurance from this company.

About the Data:
We have the data of existing Health Insurance Customers, this Data includes 12 relevant data points such as age, gender, sales channel data, vehicle ownership data.
And most importantly, the target variable: whether the customer has vehicle insurance or not. The Data is available for 390K existing customers.

Approach taken:
The task was divided into 2 main parts:
1.**Statistical Analysis** over the dataset to discover relationships between each feature and the target variable. So that this relationship information can be used by the management in making better Business decisions
2.**Creating a Machine Learning Pipeline,** that can take in the data of any new customer and predict whether they will be interested in vehicle insurance. It was required to keep this pipeline modular, such that it can be retrained often when new data is collected

Technical Details for ML: We trained 3 Different Algorithms
(Logistic Regression, Decision Tree, Random Forest)
We used **GridSearchCV** for HyperParameter Tuning
Comparing both F1 and AUC-ROC Score, we can see that Random Forrest and Random Forest model performs the best . Best AUC-ROC = 0.85, Best F1=0.43

Conclusions: Insights from exploring the Data :

- Customers of **age between 30 and 55** are **more likely** to buy insurance.
- Customers with Driving License have higher chance of buying Insurance.
- Customers **with Vehicle Damage** are **more likely** to buy insurance.
- Customers with **Vehicle age** between **1 and 2 years** are more likely to interested.
- Customer **who are not insured previously** are **more likely** to be interested.
- **Region 29** has given us highest conversions
- **Vehicle damage** is the most important feature that drive predictions followed by previously insured