Jordan University of Science and Technology

College of Computer Sciences & Information Technology

Computer Information Systems – CIS321 Project – Health Insurance

A project submitted

By

ASEEL ZEYAD ABDELLATIF ALHAWAMDEH 147039

JELAN AHMED ALI KHWEILEH 149491

Supervised by

Professor. Ahmad Mustafa

1. **Introduction:** Insurance companies are increasingly using machine learning to predict customer behavior and tailor their offerings accordingly. In this project, we aim to predict whether customers who previously purchased health insurance will also be interested in vehicle insurance. We use a binary classification approach to solve this problem.

2. **Data Exploration:** We started by loading and exploring the dataset. The dataset contains features such as gender, age, driving license status, region code, vehicle age, vehicle damage, annual premium, policy sales channel, vintage, and the target variable "Response" indicating interest in vehicle insurance.

3. **Model Selection:** We experimented with different classification models but found that XGBoost performed the best in terms of balanced accuracy. Therefore, we chose XGBoost as our primary model for prediction.

4. **Hyperparameter Tuning:** We used RandomizedSearchCV to search for optimal hyperparameters for the XGBoost model. The hyperparameters considered included learning rate, max depth, min child weight, gamma, subsample, colsample by tree, and the number of estimators.

5. **Model Evaluation:** We split the dataset into training and validation sets and evaluated the model's performance using balanced accuracy, classification report, and confusion matrix. The model showed promising results on the validation set, achieving high accuracy and balanced performance across different classes.

6. **Prediction and Submission:** Finally, we made predictions on the test set using the trained model and submitted the results for evaluation. The predictions were saved in a CSV file in the required format for submission.

7. **Conclusion:** In conclusion, the project successfully developed a predictive model to determine customer interest in vehicle insurance based on past health insurance purchases. The model demonstrated robust performance and could potentially assist insurance companies in targeting customers more effectively. Further refinements and optimizations could be explored to enhance the model's performance even further

We also tried another models here are some;

Model	Parameters	Kaggle score
KNN	GINI	.50
Decision tree	GINI	.65