**Professional Certificate in Machine Learning and Artificial Intelligence**

Module 25: Portfolio Project

Model Card for In-Vehicle Coupon Recommendation

**Model Details**

* The machine learning model predicts whether a customer will accept a coupon based on venue and contextual features. Gradient Boost tree is used as a classifier which gives best results among three candidates.
* Gradient boosting decision tree model has been used with low numbers of estimators 300 and tree depth = 5. Grid search is then used to tune learning rate and fraction of features used to train each tree.
* Developed by Syed Umair

**Intended Use**

* To be used by marketing agencies to identify customers for coupon on the given routes.
* Not intended for marketing policies beyond venues identified in this dataset.

**Factors**

* Dataset contains 12684 data points. Most of the data points are categorical and some are numerical.
* Data contains user attributes like age, gender, marital status, has\_children, education, occupation, income, how frequent user goes to bar/restaurants etc,. Contextual data contains weather, destination, time of driving, distance to coffee bar etc,.
* Output of the data contains 1 if the customer accepted the coupon and 0 if the customer did not accept the coupon.
* Some data is missing which is imputed. One-hot encoding is used for categorical features.

**Metrics**

* Performance metrics: ROC curve, Precision and accuracy. We are interested in how many TP and FP we have predicted by our model.
* 60% of the data is used to train the model and 40% is used to test it. Data split is stratified. 5 fold cross validation is used for model tuning.

Immagine che contiene testo, schermata, Policromia, diagramma

Il contenuto generato dall'IA potrebbe non essere corretto.

Immagine che contiene linea, testo, diagramma, Diagramma

Il contenuto generato dall'IA potrebbe non essere corretto.

**Limitations**

* Model should be used for marketing only and with the constraints that come in the dataset.

**Trade-offs**

* All three models overfit the data. However, a very large test set 40% satisfies the model performance. In future, it is advisable that model is monitored regularly for performance related issues that may arise due to model overfitting. To reduce model fitting feature dimensionality can be reduced to reduce the complexity of the model.