**Customer Insight for Tourism Enterprise**

001205751 Yiting Wang

001490068 Mingxin Liang

**A. Topic description**

Nowadays, transnational tourism become more and more popular. In order to maximize benefits, a transnational travel business wants to know about whether or not a person is interested in travel and become a potential customer. Optimize the customer experience by identifying high-value, high-potential customers and implementing differentiated customer management strategies. We plan to create a new machine learning model to realize above target. And we believe this process would be significant and meaningful to the whole tourism enterprise.

**B. Data sources**

1. We scratched this dataset from Internet. It has 17 features and 65534 rows.
2. Here is meaning of feature:

|  |  |
| --- | --- |
| **Feature** | **Meaning** |
| interested\_travel | Whether or not the customer is interested in travel |
| computer\_owner | Whether or not the customer has owned a computer |
| age | Customer’s age |
| home\_value | Property price |
| loan\_ratio | Loan ratio |
| risk\_score | Risk\_score |
| marital | Marital status |
| Interested\_sport | Whether or not the customer is interested in sport |
| HH\_grandparent | Whether the grandparents of the customer are alive |
| HH\_dieting | Customer dieting preferences |
| HH\_head\_age | Household head age |
| auto\_member | Number of customers in the club |
| interested\_golf | Whether or not the customer is interested in golf |
| interested\_gambling | Whether or not the customer is interested in gambling |
| HH\_has\_children | Whether or not the customer has children |
| HH\_adults\_num | Number of adults in the family |
| interested\_reading | Whether or not the customer is interested in reading |

**C. Algorithms**

In order to find a better model for customer insights, we implement two different models: DNN and Gradient Boosting Machine. For both these two methods, we create the base model, tune the parameter with validation dataset and evaluate model performance with the test set.

* Gradient Boosting for classification
  + Build and compile a base model
  + Parameter tuning with Grid Search
  + Implement the parameter in new model
  + Evaluate the model with test dataset
* DNN model
  + Build a DNN model with multiple layers
  + Tune parameter with the validation set
  + Evaluate the model

**D. Evaluating the success of the model**

For comparing the performance of the models we choose, we tried to evaluate the models above with the same metrics. We use sklearn.metrics tools.

* Accuracy

Accuracy classification score.

* Roc\_auc\_score

Compute Area Under the Receiver Operating Characteristic Curve (ROC AUC) from prediction scores.