

Term Project Proposal

Project Title:

ELO Merchant Category Recommendation (Help understand customer loyalty)
Kaggle Competition

Team Members: Shan Kwan Cho

Guidance By: Dr. Hui Yang

Associate Professor,
Computer Science Department, San Francisco State University

Project Description:

- About the company and the dataset:

[ELO](#) is one of the largest payment brands in Brazil, has built partnerships with merchants in order to offer promotions or discounts to cardholders. But do these promotions work for either the consumer or the merchant? Do customers enjoy their experience? Do merchants see repeat business? Personalization is key. Elo has built machine learning models to understand the most important aspects and preferences in their customers' lifecycle, from food to shopping. But so far none of them is specifically tailored for an individual or profile. ELO tries to understand the importance of aspects and preferences for their customers' lifecycle using ML strategies.

- The project will include processing using **Root Mean Square Error (RMSE)** to predict the customer loyalty. It will address the problem of how well can we predict the customer loyalty based on its 4GB size of historical_transaction.csv format. Also, it has new/old merchants.csv and identify the correlation between historical transactions and old/new merchants information.
- I am particularly interested in this problem because predicting customer loyalty using their previous and current merchant information and transaction history is essential for business development and based on prediction, the business can be improve by pointing out the weakest area and can fix effectively. Moreover, Kaggle competitions are interesting to participate and are real world business applications that actual face in the real-life.
- The data has been provide available by ELO to whoever participate in the Kaggle Competition. The data is available in CSV format and it contains information of each card historical (up to 3 months) transactions, new and old merchants information with different time.

Learning Goals:

- To learn how to analyze huge amounts of data and learn more about Root Mean Square Error (RMSE) and other necessary concepts of knowledge discovery from data (KDD).
- To be able to detect new patterns and anomalies in the data to be processed and learn more about pattern mining and cluster analysis.
- To learn how to handle huge size of data, effective data cleansing and feature engineering (which data is most relevant affecting to target)
- To expand my knowledge of Data Mining to aspects of this project I find particularly interesting.