**Identification of Churning Users** **using ML**

**The Problem**

The problem poses a supervised machine learning classification task, aiming to determine if a merchant will continue with TK or not in near future. The task may be extended from a binary classification problem (O/Ps: will get churned, will not churn) to a multi-class classification problem (O/Ps: will get churned, will do little activity, will do a heavy activity, etc.).

**Modality of Analysis**

On any given day, users of the last 42 days (6 weeks) have been analyzed. Based on their activities in the last 42 days and some additional generic features, predictions are generated in terms of whether the user will show up in the next 14 days (churning) or not (non-churning).

**How can this be helpful in business?**

Marketing can be handed over a list of churning merchants each morning based on the O/P. On the basis of this, they can design the right creatives and communications.

**Formal Method**

* Design of Sample: An initial pool of merchants is first gathered, based on active users (events + transactions) of the last 42 days. Their registration has to be before that, in order to ensure fair play.
* Design of Feature Space: 42 days have been broken into equal 7-day segments for rounded weeks. 31 features mined for the task are as follows:
  + Active days of weeks-1, 2, 3, 4, 5, 6 (6 features)
  + TRT of weeks-1, 2, 3, 4, 5, 6 (6 features)
  + TACS of weeks-1, 2, 3, 4, 5, 6 (6 features)
  + Time spent on weeks-1, 2, 3, 4, 5, 6 (6 features)
  + Roaming days of weeks-1, 2, 3, 4, 5, 6 (6 features)
  + Age of merchant in days from registration (1 feature)
* Design of Labels: Binary labels are generated so far
  + 0 (churn): if a merchant will not appear in the next 14 days
  + 1 (non-churn): if a merchant will appear in the next 14 days
* Design of Prediction Algorithm: An NN classifier has been applied for the purpose, with the following specifications:
  + 31 input features
  + A deep NN with 5 ReLU hidden layers, with **35, 25, 25, 25, 20** hidden neurons respectively
  + learning rate = **0.01**, epochs = **500**
  + SoftMax classification O/P layer
* Results:
  + Accuracy: 81% on both training and test, showing no overfitting and great generalization
  + Precision: yet to code, will be around accuracy
  + Recall: yet to code, will be around accuracy

**Probable Developments**

* Classification among multiple churn-categories
* Inclusion of features describing versions, locations, BI-types, RAU/PU-info.
* Calculation of other performance metrics: precision, recall, F1-score, etc.