Meeting Agenda with Prof. John Springer and Emily Tluangneh

Date: September 28, 2017

Time: 3:30 pm

Location: KSQ

Attendees:

|  |  |  |
| --- | --- | --- |
| Name: | Roles/Responsibilities | Present: |
| Seunggeun | Collecting strategies of obtaining data | ✓ |
| Jun | Collecting strategies of obtaining data | ✓ |
| Hoon | Collecting data from social media | ✓ |
| Nawon | Collecting data from social media | ✓ |
| Yujin | Collecting data from social media | ✓ |

New Project Ideas:

Big Data Analysis on Music Preference

-Preference on music by genres, demographics, culture and behavior

-Collect data on user’s information, components of the music, metadata, Youtube views

-KKBox has provided data

-Use Python or R code to cluster and correlate the data

-If possible, would like to participate in the KKBox’s Music Recommendation Challenge

Topic discussed with John Springer

1. Presentation given about obtaining music preferences
2. Can obtain data without actually competing
3. Split work between all the group members: pair up and work together
4. Haven’t done a lot with
5. Modeling: need more research, if numeric use support vector machine (SVM), datamining
6. People who are good at mining may not tell you what they are doing EX: Netflix or Google
7. Training and test: development of predictive models to guide many decisions
8. Problem: over trained - meaning it does so well during training that during test, it fails
9. Cross-Validation: split data into a percentage randomly to create training sets multiple time to make sure the results are matching with each other at test set
10. SVM: algorithms discriminative classifier by separating hyperplanes to make it 3D, very complicated
11. Dependency on either quantitative or categoric data, can’t be both
12. Nominal “name – label a series of values” and Ordinal “scales – order of choices”