**AP Supervisor Meeting Minutes – 17 March 2016:**

|  |  |
| --- | --- |
| Date: | 17 March 2016 |
| Time: | 1:30 p.m. |
| Venue: | SMU SIS Meeting Room 4-1 |
| Attendees: | Denise Quek Si Ying |
|  | Tan Wei Song |
|  | Prof. Kam Tin Seong |
| Agenda: | 1. Discuss methodology for revised objectives |
|  |  |
|  |  |

1. **Review of meeting’s discussion**

|  |  |  |  |
| --- | --- | --- | --- |
| **S/N** | **Issues Discussed** | **Action By** | **Due Date** |
| 1. | Approach at sieving out important variables   * Response variable, y, is a function of consumers’ social demographic patterns and shopping behavior * We can decide to split or combine social demographic variables and shopping behavior based on how we want to draw our conclusion; * By combining both social demographic and shopping behavior together in one run, we are able to see the relative importance of each variable | All to note | - |
| 2. | **Current Model Evaluation**   * Evaluate which model to use based on the accuracy of the model & complexity of the model * Our small dataset is not suitable to build a predictive model – this is evident through the instability of model results as shown in decision tree and bootstrap forest partition methods   + E.g. Decision tree: every time a split is done, the R^2 does not change much, can almost conclude that the change is insignificant   + E.g. Test dataset produces negative R^2   + E.g. Bootstrap Forest: signs of training data to be over-fitted (evident through different prediction patterns across training, validation, and test data) due to large number of predictor variables and small dataset   + E.g. Bootstrap Forest: results identified large number of variables instead of a sizeable amount – shouldn’t be that case | All to note | - |
| 3. | **Alternative Approach – Logistic Regression**   * We cannot build a predictive model anymore because of the unsuitability of the dataset. Instead, what we can do from now is to build an explanatory model. * Use logistic regression to do that because of the small sample size * Make sure variables used are statistically significant   + Evaluate based on likelihood ratio test   + Recast household income to isolate 10-15k category * Use stepwise method because we have a lot of variables   + P-value threshold 🡪 make model 🡪 run   + Because of the limitation of JMP, we have to split the variables into categories and run them separately first before combining all the significant variables identified | Denise & Wei Song | 21 March 2016 |

1. **Next meeting’s agenda**

|  |  |  |
| --- | --- | --- |
| **S/N** | **Issues Discussed** | **Action By** |
| 1. | Next supervisor meeting will be held next week. Date to be confirmed once prof’s schedule has been made available to us. By then, prepare for next week’s meeting based on the agenda stated below.  Agenda:   * Results of logistic regression | Denise, Wei Song |

The meeting was adjourned at 3:00pm. These minutes will be circulated and adopted if there are no amendments reported within the next three days.