|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Consultation | | | | | |
| 04.05.2016 | | 10:00 AM | | SIS 4.1 | |
| Meeting called by | Oh Peng Ho, Tan Jhun Boon, Yap Jessie | | | | |
| Type of meeting | Supervisor Meeting | | | | |
| Facilitator | Professor Kam | | | | |
| Note taker | Tan Jhun Boon | | | | |
| Timekeeper | Professor Kam | | | | |
| Attendees | Oh Peng Ho, Tan Jhun Boon, Professor Kam | | | | |
| Final Paper Review | | | | | |
| 60 MIN |  | | | | |
| Discussion | Presentation   * Audience in auditorium with different types of audience - Academic, Practitioners, Students * Technical and basic questions are expected * Rapidminer APAC director coming * Share experience of using rapidminer * Have to prepare poster | | | | |
|  | MBA   * Abstract   + Moderately done, acceptable. Re-edit for more impact.   + Be more explicit about the issue - importance of upselling in F&B industry   + Not many papers focus on SMEs * Screenshot receipt to put before the data   + Verbally describe transformation process, no need to show excel transformation from receipt to MBA * Template/format * Use prof’s template * Explain section for division of paper * Clear headings and subheadings * Blow up tables and diagrams * Text - single spacing * Data cleaning procedure (rapidminer) put at the data cleaning section * Weighted MBA * Weighting might need experience - might need to interview salesman * Delphi methods - expert opinions * Social network analysis (Product death spiral)   + Future * SAS EM * Depends on the focus of the paper * We can just stick with using Rapidminer and include and say that there are other alternative tools | | | | |
|  | Regression- Clustering   * Abstract   + No problem   + Highlight finding in the abstract * Data   + There’s no clear “13 hours” in the time-series of the data cube when used as for analysis   + Time-series: test for auto-correlative   + Multiple linear regression: conform to a few assumption     - Must check using VIF for **Multicollinearity,** >5 beware, >8 have correlation       * Correlation matrix       * Highly correlated, exclude them         + Ensure there’s no Multicollinearity     - Much check using **Demi Watson test** for auto-correlative on the basic model (to check whether we need to revise our model) - that mean’s it not time-independent - >2 strongly correlated       * Both X and Y can have **different** temporal patterns, but for us sales is most likely the one that is time-dependent * Model   + Since its time-series, it’s not **time-independent**   + Formula should include t: y = a + b1x1t + b1x2t * Analysis   + Past historical trend can be incorporated into analysis   + Multiple linear regression to a dynamic model   + Add in field for whether a data point is a particular hour   + Add in field for whether it’s a weekday or weekend   + For our case, the sales and productivity of the cashier is also affected by the time; if we don’t take into the account the time effect, it’d be very different analysis.   + Suggestion:     - Introduce conventional multiple linear regression     - Introduce reconstructed version of multiple linear regression - with explicit **time-factor elements** to the analysis       * We can add the time or day as a column       * Lag: data-shift the previous (absolute) sales from previous hours into another column, since there is an influence on the current sales. By adding lag as an independent variable, we provide the “cyclic” nature of the data. It’s basically to add in the effect by t-1.         + JMP has function to do this * Discussion   + Different time periods have different significance in regression   + Suggestion:     - Isolate by cashiers, and only show the times when a cashier is on duty or not - there’s no need to pool everybody in the overall model. Since we are only accessing one cashier at one point of time. If we add in too many dummy variables for the linear regression (when one cashier is working, the other fifteen will be 0), then the regression will be less complex and there won’t be as many independent variables * Clustering   + Drop clustering   + Without clustering, with the “time-factor” multiple linear regression, we have more than enough value-add | | | | |
| Action Items | | | Person Responsible | | Deadline |
| Regression Paper | | | All | | 17 Apr 2016 |
| MBA Paper – abstract and formatting | | | All | | 17 Apr 2016 |
| Posters | | | All | | 17 Apr 2016 |