|  |  |
| --- | --- |
| Date/Time | 20th Mar 2017, 4:00pm to 5:00pm |
| Venue | SMU SISS Meeting Room 4.5 |
| Attendees | Prof Kam, Albert, Jun Liang, Russell |
| Agenda | 1. Share and get feedback for materials to be presented to sponsor in upcoming sponsor meeting
2. Signing of revised version of NDA
 |

Consultation Meeting Minutes:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Task/Description | Person in Charge | Due Date |
| 1. | Signing of NDANDA was finally agreed upon by all parties and signed by all.  | All | 20 Mar 2017 |
| 2. | Sharing of material to be presented to sponsor – Pivot TablesProfessor Kam suggested using a bullet table and also to use more summary statistics to better give a benchmark like the median, upper quartile, lower quartile etc. These would allow the user to easily compare results at a glance. Professor Kam suggested the use of Sparklines/sparkTable in R (<https://journal.r-project.org/archive/2015-1/templ-kowarik-meindl.pdf>) to better display results as well as to expose us to analytics software.Professor Kam commented that we should be more selective on the indicators of performance. Created variables like word count or sentence count may not be relevant to reflect performance. He suggested reviewing literature that discuss how performance of web posts are gauged.  | All | 23 Mar 2017 |
| 3. | Sharing of material to be presented to sponsor – Articles regression modelProfessor Kam commented that we must be cautious in using the sentiment score dictionary because it may not be within the same context as our sponsor – there needs to be some contextualization done if it is to be effectively utilised.Professor Kam also suggested the use of a flow diagram to illustrate how the sentiment analysis is done. Regarding the sentiment analysis, Professor Kam suggested taking samples of articles with sentiment score analysis done to review if the score reflects the actual sentiment of these articles; he also suggested checking whether a large proportion of the words have been tokenized or flagged out from the articles to gauge the effectiveness of the sentiment analysis. Regarding the regression model, Professor Kam clarified the steps that he would use for analysis. The first step would be to check the Variance Inflation Factors (VIF) for continuous variables to remove multicollinearity from the model. The second step would be to do a simple linear regression to see how linear each independent variable is. For predictor variables, transformation can be done if needed (based on distribution analysis). For response variables, exclude outliers (or transformation too if needed). Professor Kam also suggested that each topic for articles can be recoded as dummy variables. Professor Kam explained how to read the variable for authors within JMP Pro 13. He said that the concatenated variable would show who ‘contributes’ to the fit of the model, meaning that whoever is inside would contribute. After the ‘-‘ sign, these people don’t contribute to the fit of the model.  | All  | 22 Feb 2017 |