

Meeting minutes

- Team SkyTrek

Date (dd.mm.yyyy)	22.03.2016
Attendee(s)	Jedaiah, Aseem, Viet Huy, Prof Kam
Absentee(s)	NA

Contents

Item	Discussion point	Remark
1	Points to include in Final Report	<p>Should talk about the initial impression of the project vs that after receiving the data from the client</p> <p>Transparency in process of preparation of data → reviewing of analytical methods to be done</p> <p>No need to test accuracy of clustering. Just be able to describe the clusters obtained.</p> <p>Include details about the exploration phase. Record all the models that were explored. Hence, evaluate the models before choosing the final model</p> <p>What is the basis for binning the data .e.g. equal interval vs Quanta bin Instead of artificial binning → run decision tree to see the cutoff point of binning</p> <p>When build model → include all predictors and let the model exclude them</p> <p>Final report should focus on method of analysis, model assessment → reflect on why some model perform better than the others.</p> <p>Talk of data preparation, but giving more focus on the analysis and explain the method. Talk about how the algorithm works.</p> <p>Provide more details on the use of RapidMiner as a tool itself.</p>

Actions

Task	Assigned to	Status
Update mid term report with technical analysis	Team	Done
Explore details of incorporating categorical and numerical data into regression <i>model</i>	Huy	Done
Get Skyscanner NDA to Prof Kam	Aseem	Done
Update dataset with content theme allocation for each URL	Jedaiah	Done
Analyze text mining clusters for titles	Jedaiah	Done
Run regression analysis again	Aseem	Done
Add article titles to dataset	Huy	Done
Wiki update for week 8	Aseem & Jedaiah	Done
Manual classification of content types based on heavy weightage terms generated from k means model	Jedaiah	Done
Book consultation slot with Prof Kam for week 8 Tuesday this Sunday <ol style="list-style-type: none"> To clarify with Prof Kam on issues mentioned in proposal feedback How should we account for the NDA since the final conference paper would eventually be published online?	Jedaiah	Done
Complete wiki updates for week 7 <ol style="list-style-type: none"> Minutes clean up and upload Updated Gantt chart	Huy	Done
Complete data pull of 7 datasets and get them verified by Aseem Remarks: Huy will be out of town for week 8 and must ensure that pulled dataset is accurate and ready for analysis in week 8	Huy	Done
Pull Google Analytics Data	Aseem	Done
De-aggregate shares count for each URL	Jedaiah & Huy	Done
Build regression model based on current dataset	Aseem	Done
Come up with list of possible title categories to clear with sponsor	All	Done
Clarify with Prof Kam: Is there a way to dynamically assign new articles to the relevant clusters based on the trained model?	All	Done
Characteristics to scrape from article: <ol style="list-style-type: none"> Number of words (remove stopwords) Number of links Images Videos Number of shares of article (by Jed) 	Huy	Done
Retrieve URL share count	Jedaiah	Done
Put topic modelling and google trends research and implementation into Gantt chart timeline	Jedaiah	Done
Input meeting with client in gantt chart for: <ol style="list-style-type: none"> 25th Jan 2nd Feb 11th Feb 	Jedaiah	Done
Create flow chart documenting process how skyscanner team plans for the next quarter's content postings. To send it to client for verification	Aseem	Done

Test Google API public holiday pulling	Huy	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> To plot Unique page views views (count) vs Week of article publishing Clicking this particular bar of interest (week has high count) would drill into the next bar chart (Number of unique page views, vs identifying URL) Further clicking on the URL would reveal characteristics of the article (e.g. number of images, videos, etc, bounce/exit rate, etc) 	Jedaiah	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> Identify key features of popular posts (MLR against unique page views) 	Huy	Done
Create a mockup for the following visualization: <ul style="list-style-type: none"> Where do they come from (On the whole country news page level, where do the traffic come from) How long do they spend at the page (average time on page) Do they visit other pages? (bounce rate vs exit rate) How effective are the recommended pages? (If GA does not track the unique page views, then this analysis is not going to be implemented) [Just to prove a logical point] Correlation test between number of shares vs unique page views 	Jedaiah	Done
Deal with proposal segments: <ol style="list-style-type: none"> Sponsor and Background Information Motivation Objectives Data Scope of Work 	Aseem	Done
Questions for Prof Kam: <ol style="list-style-type: none"> Ask Kam about requirement to post sample date. Concern: data confidentiality What is required for scope of work? (different from methodology) 	All	Done

The minutes were prepared by: Viet Huy and Jedaiah Tan