BiZiT Festival 2017 - Big Data Challenge 1

Overview:

As Singapore's leading infocomm provider, Singtel is committed to continuous innovation and service improvements to meet the evolving needs of its customers in today's digital world. This is because the way consumers make their buying decisions and choose to engage with the telco has changed along with their communication and media consumption habits.

Recognizing the trend where a growing number of customers prefer the convenience and immediacy of self-service and online channels, Singtel has adopted an omni-channel customer service approach through deepening its digital engagement efforts via My Singtel App, live webchat and social media channels such as Facebook and Instagram.

Specifically, Singtel has dedicated staff and customer care officers to manage and to respond to any queries or feedback posted on the Singtel Facebook page. With the increasing number of Facebook posts related to Singtel, it is sometimes a challenge for Singtel staff to browse through all the posts and to respond adequately in a timely fashion.

Case Challenge Statement:

For this competition, you are expected to propose how you can make use of the data from Singtel's Facebook page to enhance its customer care's effectiveness and efficiency.

For instance, you may propose to build a model from the Singtel Facebook posts training dataset to automatically classify future customer feedback and call record cases, based on the Singtel case classification taxonomy provided.

You are expected to illustrate the business value of each of your proposed data analysis.

You will be provided with the following data should your team proceed to semi-final round:

1. Raw data of Singtel's Facebook public page posts (i.e. include post date and post author information)
2. Singtel case classification taxonomy (i.e. a list of case categories)

You may propose to include any additional open-source datasets that may make enrich your data analysis and results.

Evaluation:

Evaluation during the preliminary judging stage will be based on the quality of the data analysis solution presented in each team's proposal.

Evaluation during the semi-final and final stages will be based on both the quality of the actual data analysis and the quality of the business solution recommendations.

For instance, if the proposed solution include case classification, the appropriate and applicable metric to apply will be the Macro F1-score (MaF) to evaluate the classification accuracy on the final test dataset.

The MaF is calculated as follows:

$$MaF = \frac{2 * MaP * MaR}{MaP + MaR}$$

For a set of classes $C = \{c_1, \dots, c_k\}$ macro precision and macro recall are calculated as follows:

$$MaP = \frac{\sum_{i=1}^{|C|} \frac{tp_{c_i}}{tp_{c_i} + fp_{c_i}}}{|C|}$$

$$MaR = \frac{\sum_{i=1}^{|C|} \frac{tp_{c_i}}{tp_{c_i} + fn_{c_i}}}{|C|}$$

where tp_{c_i}, fp_{c_i} and fn_{c_i} are the true positives, false positives and false negatives respectively for class c_i.

The organizer will review and advise on the appropriate evaluation metrics according to each team's proposed solution.