Supervised Classification with Full Spectrum or Premium Subset? **(Group 7)**

1601 Introduction

1603 dataset description (show improvement possible from other speaker)

1608 feature selection (show improvement possible from other speaker)

1611 Modelling (logistic, linear discriminant analysis, random forest)

1613 Results

1616 Analysis (show possible improvement selection, parameter tuning)

• Summary of the report.

The report investigates the underserved population and help them has a positive loan experience, Home Credit makes use of a variety of alternative data to predict their clients' repayment abilities. Raw datasets are preprocessed and model building with diagnosis followed by selection and exploratory analysis. LDA is the best model with 0.770 score on Kaggle.

• Strengths of the report.

Strong supportive data visualization (bar chart, pie chart and heatmap) for the data analysis section.

Couple of models used (Logistic Regression, LDA, random forest)

Analytic skills are shown in model comparison comparing the ROC curve   as support

• Describe the weaknesses of the report.

Overall, the report is well designed. Unfortunately, hyperparameter settings did not mention in the report.

• Evaluation on quality of writing (1-5): Is the report clearly written? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the

clarity of the paper, and provide details of typos.

5

The report is written in a neat way with subtopic, objectives and methodology are stated clearly. Lots of figures for visualization and demonstration. There is no type in report. I would like to suggest that the details of model configurations should be explained.

• Evaluation on presentation (1-5): Is the presentation clear and well organized? Are the

4 The presentation is clear and organized. The time allocation is balanced which divided presentation into 3 parts (data model and analysis). However, the flow of presentation sometimes is not steady due to thinking or stuck in some points. The presentation explains models with graphs and mathematical result.

It shows improvement needed which is good.

• Evaluation on creativity (1-5): Does the work propose any genuinely new ideas? Is this a work that you are eager to read and cite? Does it contain some state-of-the-art results? As a reviewer you should try to assess whether the ideas are truly new and creative. Novel combinations, adaptations or extensions of existing ideas are also valuable.

4 The model part is good for reader have easy understanding and may even inspire the reader, speaker provided lots of information about the model used and the idea behind the selection. However, the preprocessing of dataset is not provided. The reader has idea how to re-construction the model but may need multiple of try for same result and manipulate the result. That would be a great presentation to reader and reader may cite this report.

• Confidence on your assessment

3, I have carefully read the paper and checked the results