GSEM, University of Geneva   Spring Semester 2020
Machine Learning ('S403011')   **Report 1 assessment — Group: ML_B**
Prof. Sebastian Engelke   Assistants: Nicola Gnecco, Cesare Miglioli

General remarks

1. Define clearly the notation (e.g. y response variable etc.) in the introduction. Then you will use it in the model part to describe briefly each method.

2. A report is a self-contained, organized and clear document. It is not a collection of R outputs.

3. It is necessary to clarify the type of cross-validation scheme you use. Moreover you should provide, in the results section, at least both the test set accuracy and the leaderboard accuracy that you obtain on Kaggle.

4. Consider rounding any digit up to a fixed level that you choose at the beginning and you follow for all the document.

Individual remarks

1. Consider to not show the accuracy results (table 1) directly in the exploratory data analysis as this may confuse the reader. Place them in the right position.

2. Models section: to improve your understanding of each method you can do a sensititity analysis on the hyper-parameters. For example for random forests, you can show a plot that has different values of *mtry* and see how the accuracy changes (sensitivity) as the hyper-parameter changes.

3. Present the outputs in the Annex in a more organised way to avoid unclear messages (e.g. all confusion matrices one after the other is not useful).

Grade

2. To improve
Follow the comments above and you will greatly improve the quality of the report which has potential.