Peer Review Group 9

1. Summary
   a) In terms of the data preprocessing, group 9 not only used the basic application_train csv but also 6 additional sub-tables to enrich the dataset and features. Besides, columns with missing values more than 90% are removed. Categorical features are encoded with one-hot method, while numerical features are exacted with some point estimation method such as mean and sum.
   b) About the models, they chose logistic regression as the baseline, and after that use LightGBM to achieve high accuracy.
2. Strength and Weakness
   a) Strength
      i. Use LightGBM to achieve parallelism, high-performance and do not need huge memory any more.
      ii. Use 6 more sub-tables to enrich the dataset and features to make the model more convincing and reliable
      iii. Compare the influence on different feature combinations on the results
   b) Weakness
      i. Do not show the details of the hyper-parameters selection
      ii. Though LightGBM is good, there are limited models used
      iii. More figures and examples will be better
3. Clarity and quality of writing：4

   The report is clearly written with examples and figures. It is well organized and readily comprehensible without typo and grammatical mistakes. One area that could be improved is adding more figures and experiment results.
4. Technical quality：4
   a) The feature engineering as well as the model selection reasoning are both sound and reasonable. It is a good idea to use LightGBM to deal with large data to increase efficiency.
   b) However, they can use more evaluation metrics such as F1 scores because the dataset is heavily skewed and it is better to discuss and compare more models instead of only using one tree-based model and the baseline.
5. Overall rating：4
6. Confidence：3