1. **Full name (as in NRIC) and email address (stated in your application form).**

Name: TONG Shanshan,

Email. [tongshanshan117@gmail.com](mailto:tongshanshan117@gmail.com)

1. **Overview of the submitted folder and the folder structure.**
2. **Instructions for executing the pipeline and modifying any parameters.**
3. **Description of logical steps/flow of the pipeline. If you find it useful, please feel free to include suitable visualization aids (eg, flow charts) within the README.**

click the ‘run.sh’ file to run the .py files

In case the run.sh file doesn’t run, I put all the codes in src.ipynb under the same folder.

1. **Overview of key findings from the EDA conducted in Task 1 and the choices made in the pipeline based on these findings, particularly any feature engineering. Please keep the details of the EDA in the `.ipynb`. The information in the `README.md` should be a quick summary of the details from `.ipynb`.**

The key findings from the EDA include:

1. 17% of club members opt for attrition.
2. On average, members who opt for attrition spend 2 hours less than those who keep membership.
3. Thomson Branch with most members has the lowest attrition rate of 15%, which is 5% less than other two branches (Changi and Krangi). For female members, Thomson has the lowest (12%) attrition rate while Krangi has the highest (30%).
4. Members with travel time > 45 mins show a higher Attrition Rate (21%) than those with Travel time < 45 mins (15%-16%).
5. Members holding higher qualifications and lower tier membership tends to have lower attrition rates. This trend is more evident among female members.
6. **Explanation of your choice of models for each machine learning task.**

4 models are selected to predict attrition, they are: Logistic Model, Random Forest Model, LightGBM Model and SGDClassifier Model.

1. **Evaluation of the models developed. Any metrics used in the evaluation should also be**

**explained.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Targets | 0 | | | 1 | | |  |
| Model | Precision | Recall | f1-score | Precision | Recall | f1-score | Accuracy |
| Logistic | 0.84 | 0.71 | 0.77 | 0.22 | 0.38 | 0.28 | 0.65 |
| Random Forest | 0.85 | 0.92 | 0.88 | 0.38 | 0.23 | 0.29 | 0.80 |
| LightGBM | 0.84 | 0.93 | 0.89 | 0.39 | 0.20 | 0.22 | 0.80 |
| SGDClassifier | 0.84 | 0.93 | 0.89 | 0.39 | 0.20 | 0.27 | 0.80 |

1. **Other considerations for deploying the models developed.**