

3. Group 7: “Pawpularity” Prediction Using Meching Learning with Tabular Metadata and Images (SHAO Zhihao)

a. Summary of the report

The topic of this group is about the Pawpularity Prediction. Pawpularity is a composite metric designed by Petfinder.my organization, which measures the popularity of pet photos, estimated using traffic data across the web. He included exploratory data analysis of the data, experimented with simple models, and built complicated models.

b. Describe the strengths of the report

In the project, random forest, light GBM, voting regression and swin-transformer are used for prediction.

c. Describe the weaknesses of the report

After choosing the prediction model, this group did not attempt to improve the accuracy by tuning the hyperparameter in the models. I believe parameter-tuning is crucial if one wants to achieve the highest prediction accuracy.

Also, there is not much conclusion drawn from using different models for prediction. It would be better if this group includes why and how the models behave differently and result in similar accuracy based on their algorithms.

d. Evaluation on quality of writing (1-5): 2

The poster uses clear language. However, the amount is content and details about models are limited. It would be better to incorporate more data insights for the users such as adding the feature importance graph in the poster to better present the result of modelling. Also, the group could include more details about swin-transformer like how is it suitable for this particular project the group is working on, and what its comparative advantage over other models are.

e. Evaluation on presentation (1-5): 2

Most content in the presentation are included in the poster and the source codes, and there is not much specifically about image prediction in the project (only the swin-transformer model)

f. Evaluation on creativity (1-5): 1

Since this is a starter project in machine learning, many techniques have been covered in others' works. The writer just included several data pre-processing techniques, and common machine learning models. It was expected that more variety of the machine

learning techniques can be demonstrated, like ensembling methods or state-of-art deep learning models.

g. **Confidence on your assessment (1-3): 3**

I have carefully read the paper and checked the results