Milestone M2 report

Current State of the Project:

During this milestone, our team has made significant progress in the ML component of the project. We built upon the previously collated dataset of game metadata and images to train the title detector model and the game recognition model. We had begun working on this model towards the end of the previous Milestone and so its technicalities are described in the Milestone 1 report. After exploring with various training techniques and carefully fine-tuning, the model now performs quite well at recognizing user input pictures of the game covers. Our model achieves a test accuracy of around 90% and can recognize around 400 popular game titles with ease.

In addition to this, our team has also worked on the game recommender model. We realized that in addition to the game metadata dataset (games_info.csv) used by the image recognition/classification model, we required another dataset (user_ratings.csv) that contains individual user ratings per game for training our recommender model. Hence, we worked together to create such a dataset by scraping Metacritic pages for the (username, rating) pairs and manually entered the information for titles that could not be parsed automatically. After the additional dataset is ready, we started building and training our recommender model. The model is built upon the idea of collaborative filtering. We used two embedding in the model, one for the users and one for the games. After training this model on the user ratings dataset we collected, we have achieved an acceptable test loss of around 2 by applying techniques such as data augmentation. Also, in terms of actual recommendations, the model produces reasonable results for many tested inputs. Sample outputs are included in 'model_building_and_training. ipynb'

Finally, just as was planned in the proposal, we have begun research on app development with Android Studio. We have begun working on the frontend of our app, and we now have a skeleton of our app's UI ready (in the 'GameRecommender' directory). The fully completed version of the app's frontend will be submitted as the deliverable for Milestone 3.

Feature Changes to the Proposal:

N/A. We did not hit any dead ends/roadblocks that would require us to change any features from the proposal during this Milestone.

Current Challenges/Bottlenecks:

- With regards to our image recognition model, it works extremely well for images taken from Google images. However, when we input images that we have taken using our phones, the lighting/glare causes some issues. We believe that with further augmentation and training, this issue can be overcome.

- With regards to our game recommender model, the initial user ratings dataset is very sparse, with around 97000 entries and 60000 unique users. We have tried data augmentation on this dataset, and it has greatly improved the test loss. However, for some of the tested inputs, the model is not producing very

reasonable result when it comes to actual recommendations. We think this problem can be solved by further augmenting the dataset carefully(to make sure that we don't introduce some sort of bias) and also by using more sophisticated method for making decision about which games to recommend given the model outputs.

Team Member Contributions:

- Aayush: Trained and tested the image recognition and classification model which is now working with an acceptable accuracy.

- Harshil: Manual entry of Metacritic pages for the additional dataset, worked on building and training the recommender model and wrote the M2 report.

- Kolton: Manual entry of Metacritic pages for the additional dataset and worked on the skeleton UI for the app.

- Nick: Collating the dataset of user ratings by web-scraping Metacritic pages, and worked on building and training the recommender model.