Group Contribution Document

Mini Project

Date: 31–05–2024
Group name: DE-II_10
Group members: Sahil Sandal, Yagna Karthik Vaka, Erik Larsson, Anand Mathew Muthukulam Simon, MD Masudul Islam

Contributions

Sahil Sandal: I contributed with all the group members to project selection and planning. My main task involved preprocessing the dataset, training the dataset using various regression models, conducting hyperparameter tuning with Ray Tune using the random search algorithm and ASHAScheduler, and evaluating model performance to identify the most accurate model for predicting GitHub repository stars. This resulted in finding the best training model achieving the highest R-squared value. Additionally, I contributed to the preparation and writing of the presentation and report.

Yagna Karthik Vaka: Aggregated and collected the right features of the whole main dataset for the project using GitHub APIs and also contributed to developing the models using grid search in Ray tune and found out better accuracies with the main Test set and took the data cleaning and visualization part in the data analysis for the dataset and presented some meaningful Plots during the presentation. Joined the project meetings and collaborated on planning with team members. Finally, I contributed to the report and proofreading.

Erik Larsson: Developed the project plan and helped with designing the system architecture. I configured the client, production and developer server and made the ansible playbook to automate all the configurations as well as setting up the git hook automatically from the playbook, wrote a part of the discussion and made graphs for our results, also took part in the presentation and helped making the slides,

Anand Mathew M S: Developed the project plan with the group and designed the system architecture. I used PyGithub to fetch the preliminary dataset, configured Docker for the production and development servers, testing and troubleshooting issues with Ray cluster on development server, developed script to choose the best model based on accuracy scores and move it to the directory for git hook, remodeled dashboard and set it to load the best model and display results sorted on the predictions. Additionally, I took part in the presentation, documented the repository and contributed to the report.

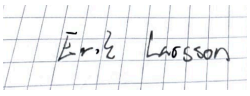
Md Masudul Islam: Collaborated with group members to develop the project plan. Worked on model training and the data engineering part, fine-tuning the code to achieve the best performance. In the report, I contributed to all parts and critically analyzed the results and scalability analysis sections. Additionally, I provided detailed explanations of the methodology, including GitHub repositories, and Docker for containerization. Summarized performance results

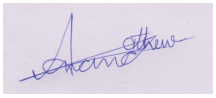
and scalability analysis, highlighting the effectiveness of ensemble methods in predicting GitHub repository stars.

Signature and Date:

Sahil Sandal 31.05.2024

Yagna Karthik Vaka Yagna Karthik Vaka 31.05.2024

 Erik Larsson 31.05.2024

 Anand Mathew M S 31.05.2024

 Md Masudul Islam 31.05.2024
