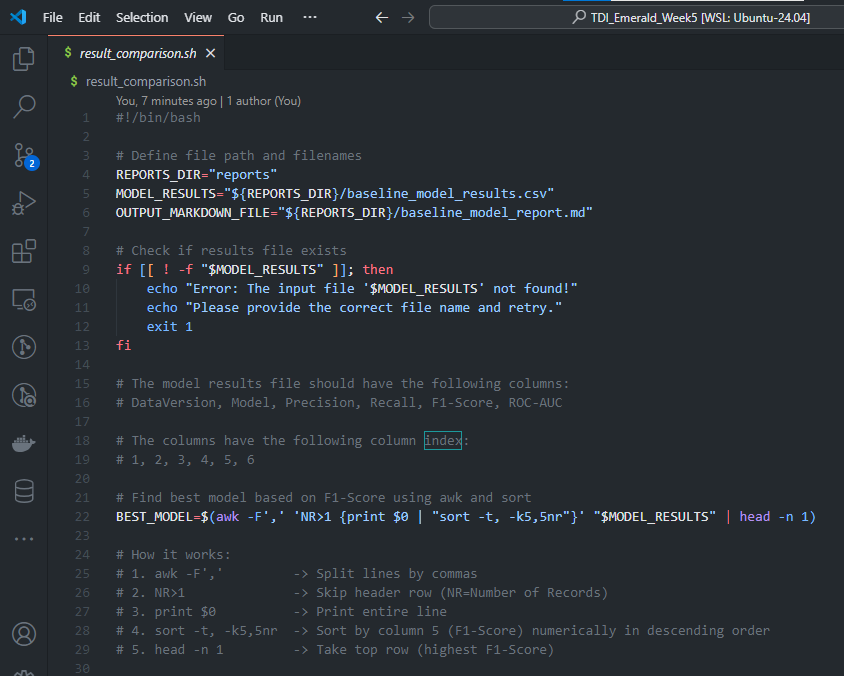
WEEK 5: Bash Scripting & Git

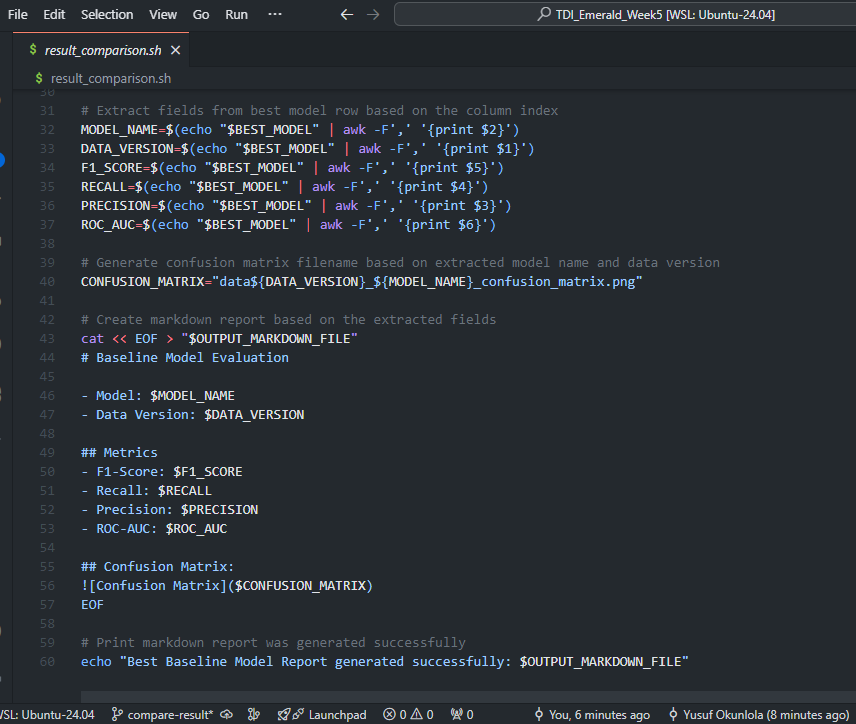
# Shell Scripting and Baseline Model Selection

Assume you successfully completed a machine learning project with the goal of selecting a baseline model based on multiple data versions. The project was originally executed in a Jupyter notebook, but now that you want to deploy it to production or cloud services, you must structure it in organised folders and files. This structure will have a report directory with model performance metrics and outputs.

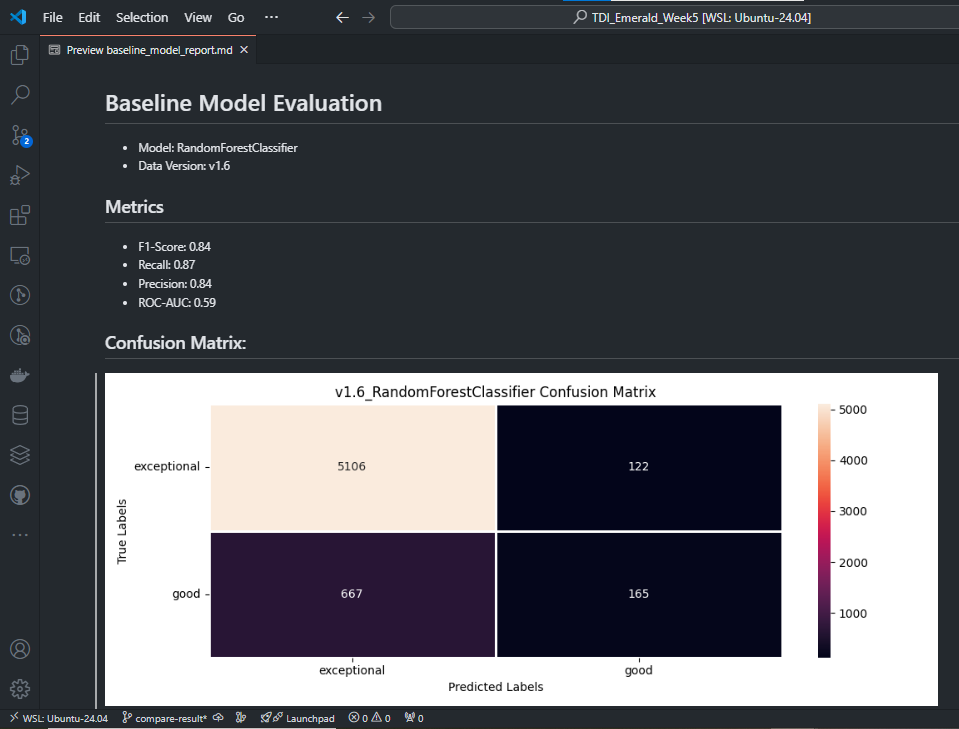
1. Create a Bash script named result\_comparison.sh in the repository’s root directory.

1. This script should automate and identify the best baseline model by:
2. Iterating *baseline\_model\_results.csv* in the report directory.
3. Extract and filter relevant best model by F1-score metric.



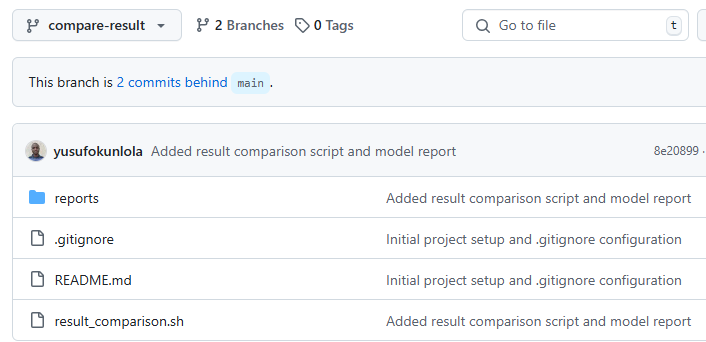


1. Generating a Markdown file baseline\_model\_report.md that summarizes:
   * Data version
   * Model name
   * Key performance metrics (F1-score, accuracy, etc.)
   * Confusion matrix image file (can be found through version number and model name)

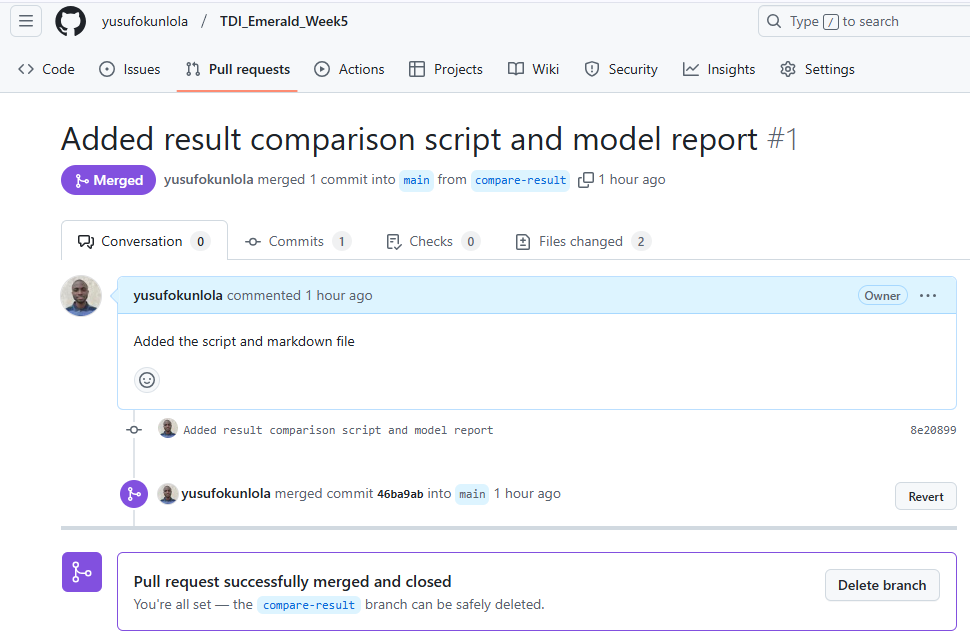


1. Once baseline\_model\_report.md is generated, add, commit, and push the file to the *compare-result* branch.

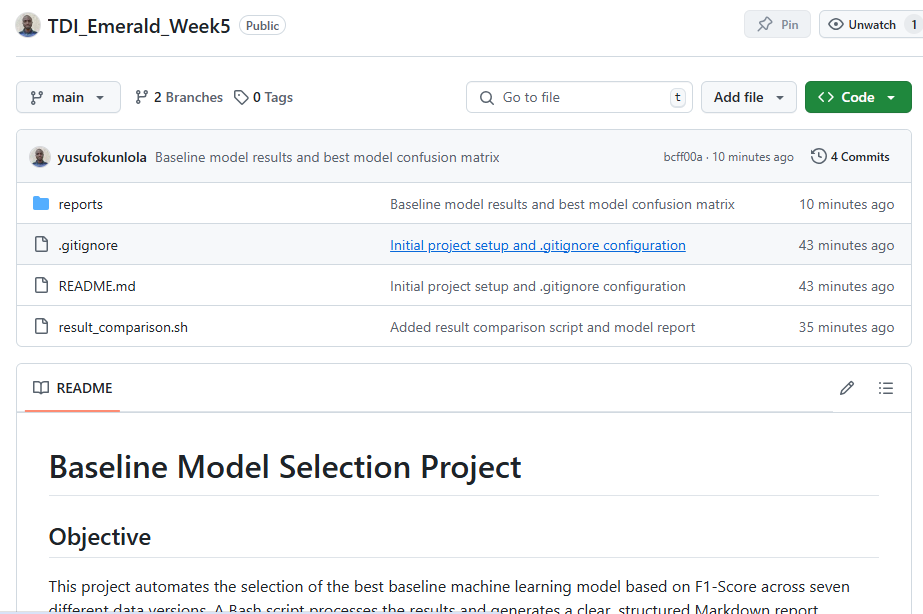
**GitHub Repo**: <https://github.com/yusufokunlola/TDI_Emerald_Week5>

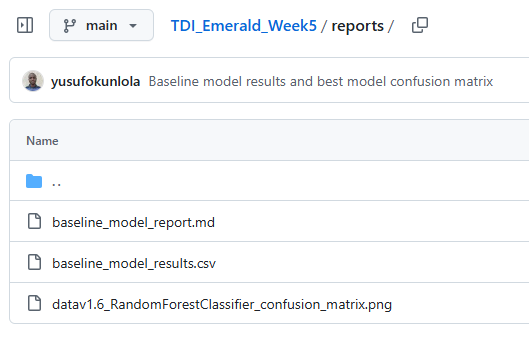


1. Submit a pull request to merge compare-result into the main branch.



1. In the main branch, ensure that only the best model report is included in baseline\_model\_report.md as per the format shown in the provided screenshot in the assignment folder.
2. The final repository in the main branch should also include:
   * The confusion matrix image of the best model.
   * The full results in CSV format ‘*baseline\_model\_results.csv’* , with all metrics for reference.

******

******

***Reference***

**GitHub Repo**: <https://github.com/yusufokunlola/TDI_Emerald_Week5>