

PREDICTION OF PROCESS PERFORMANCE USING ML

Jiaqian Huang

INTRODUCTION & MOTIVATION

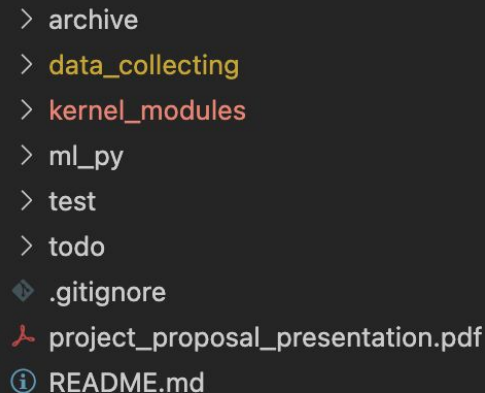
1. Benefits to predict the process performance
 - a. Early detection of performance issues
 - b. Improved resource utilization
 - c. Better capacity planning
 - d. Reduced downtime
 - e. Improved user experience
 - f. Cost savings
2. Goal: apply Machine Learning (ML) techniques to learn the performance and behavior of processes in operating system

BRIEF APPROACHES AND IDEAS

1. Collect data on both Linux and Macbook
2. Preprocess data
3. Train ML models and predict the performance

MY PROGRESS

1. Project design - **Done**
2. Kernel module in C - **Done with Demo**
3. Data collecting on Linux in C - **Done with Demo**
4. Data collecting on Macbook in Python - **Done with Demo**
5. Data preprocessing in Python - **Done**
6. ML Models in Python - **WIP**
7. Test cases in either Python or C - **WIP**

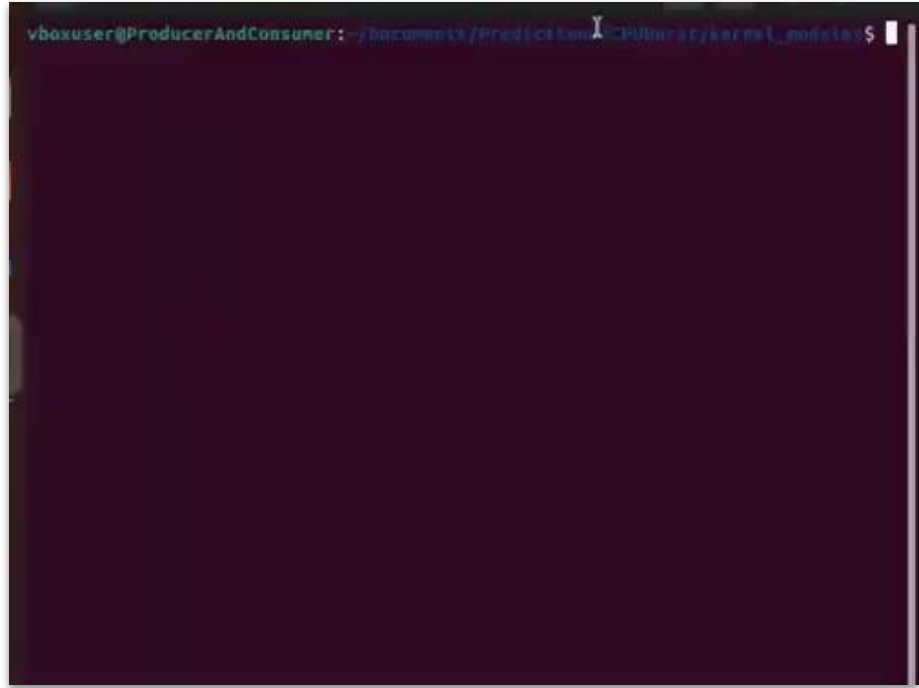


A screenshot of a file explorer window with a dark background. It shows a list of files and folders. The first six items are folders, each preceded by a right-pointing chevron (>). The next three items are files, each preceded by a small icon: a diamond for .gitignore, a red bookmark for project_proposal_presentation.pdf, and a blue circle with an 'i' for README.md.

- > archive
- > data_collecting
- > kernel_modules
- > ml_py
- > test
- > todo
- ◆ .gitignore
- 🔖 project_proposal_presentation.pdf
- ℹ README.md

DETAILS TO APPROACH

1. Build a kernel module to log custom process data in C language and install it on Linux



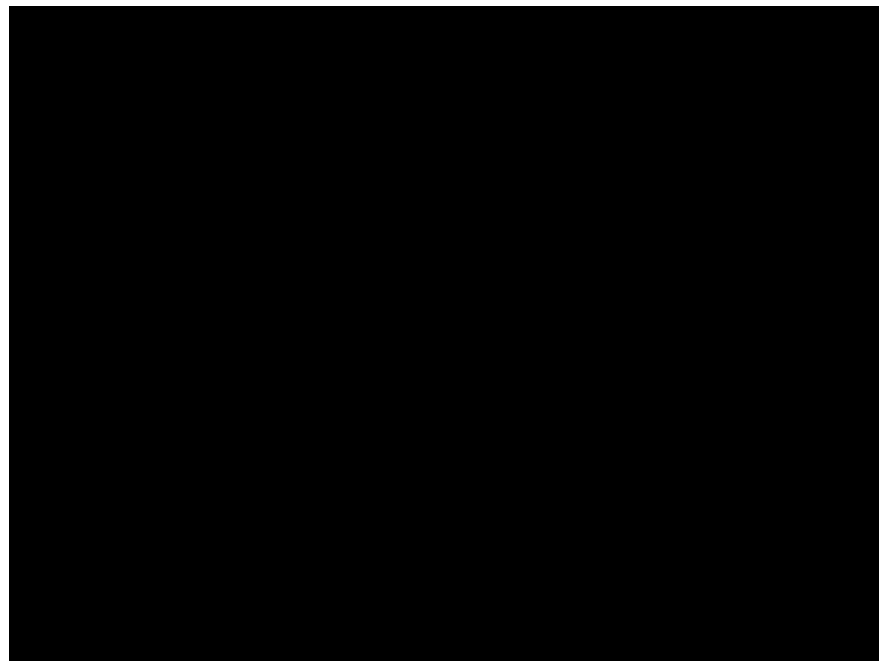
DETAILS TO APPROACH

1. Write a C file to export data from virtual log file into actual data file
2. Code a python function to collect data on macbook
3. Answer critical thinking from project proposal: Is the prediction made by the data I collect general enough? **Yes. We have collected both Linux data and Mac data which represents general users like us.**



STEPS TO APPROACH

1. Write a data preprocessing Python function to drop unnecessary data and fill the missing data in rows
2. Build ML models with help of ``sklearn`` in Python
3. Answer critical thinking from project proposal: Does our approach make it simple to predict the performance in ML? **Yes.**
Sklearn library is able to provide some types ML models for prediction. Besides, there are some tutorials online for reference.



CHALLENGES & FUTURE STEPS

1. Limited resources and various restriction on building Kernel module
2. Some data not available such as io counter and net counter when I use python

1. Modify and wrap up the ML prediction
2. Create more test cases
3. Answer critical thinking from project proposal: Does our model provide accurate prediction of performance? **WIP to build test cases to verify.**

Q&A / FEEDBACK