



**Knight Foundation
School of Computing
and Information Sciences**

Knight Foundation School of Computing and Informational Sciences

Summer 2022 Capstone II Project

PluMA 2.0 Machine Learning Pipeline Refactor



Students John Sehuwerert
Mentors: Professor Trevor Cickovski, Product Owner
Instructor/Faculty: Dr. Masoud Sadjadi, Florida International University

PROBLEM

The software pipeline is a standard analysis tool in bioinformatics. "Pipeline" is a generic term for any sequential software control flow, where the output of one program feeds as input to the next. The inherently interdisciplinary bioinformatics landscape currently contains a great deal of pipelines, often implemented as single large scripts or functions. This takes away opportunities for bioinformaticians to reuse internal stages of these pipelines, often resulting in time reinventing the wheel that could be more optimally used to advance research.

CURRENT SYSTEM

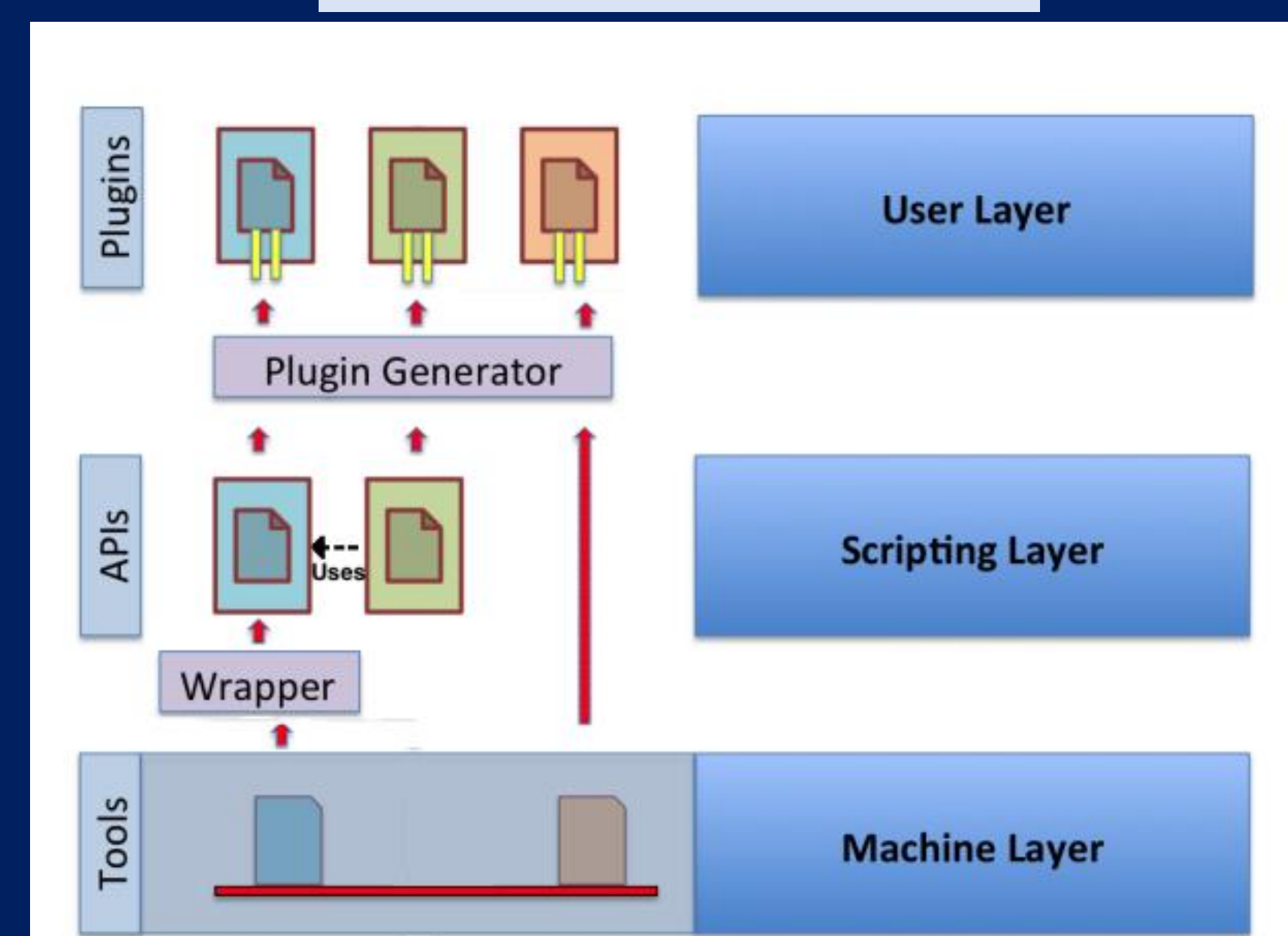
The current version of the PluMA application :

- Offers a variety of plugins for various uses in a wide range of applications.
- Can execute stages of a software pipeline through the use of dynamically loaded plugins.
- Supports plugin creation for a wide variety of programming languages (Python, C++, etc.)

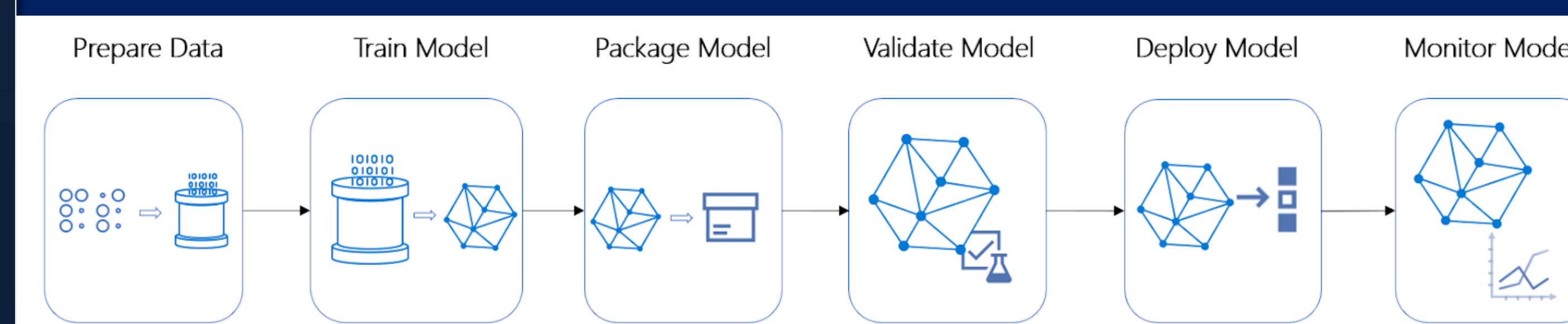
REQUIREMENTS

The goal of this project will be to properly refactor these pipelines as a series of plugins using software design patterns, each of which will then be redistributed open-source and usable in other pipelines, enhancing the impact of the work of our collaborators and facilitating community-based development.

SYSTEM DESIGN



OBJECT DESIGN



IMPLEMENTATION



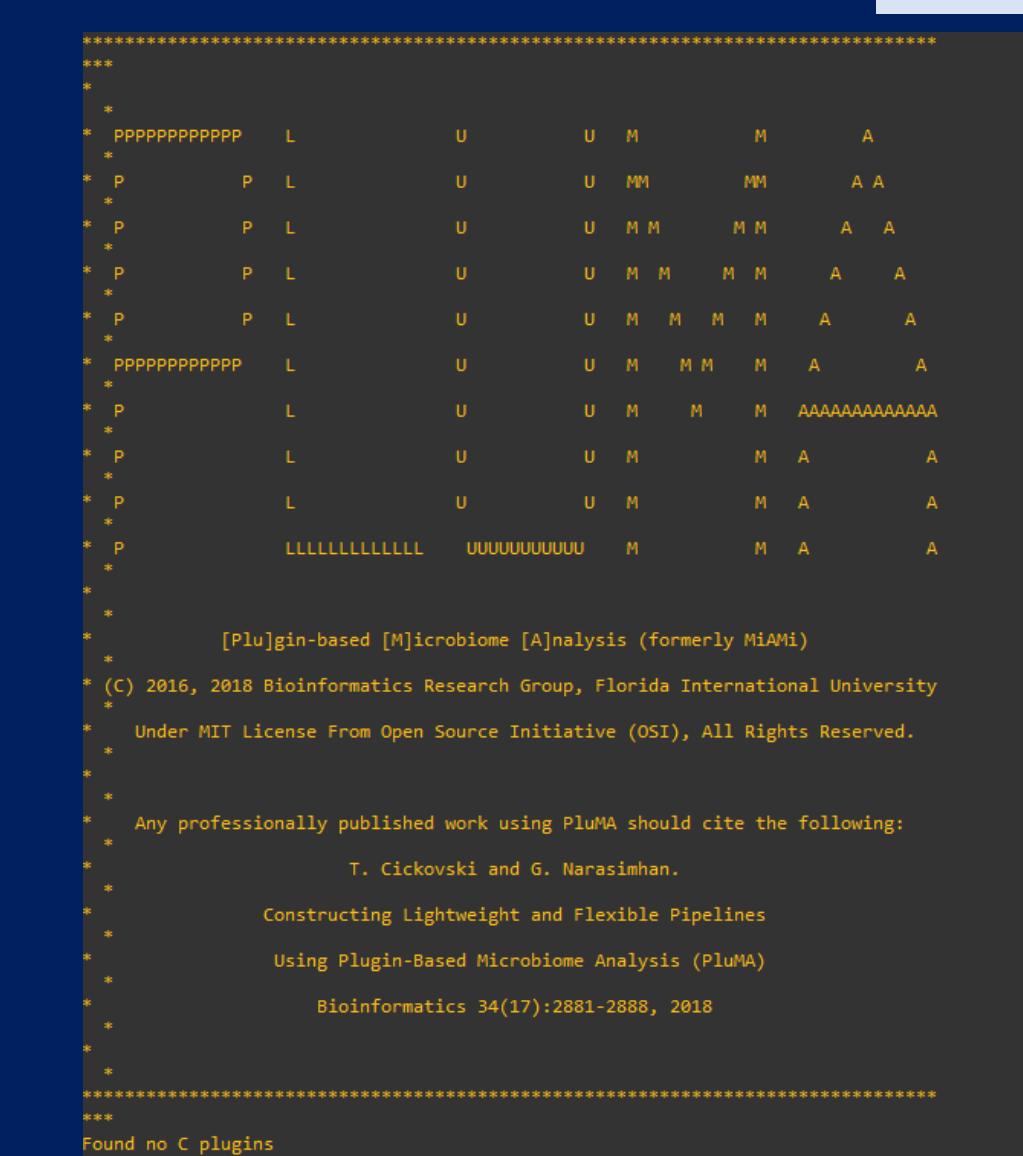
VERIFICATION

Prior to merging with the main codebase, the team was split into teams of two to tackle each implementation and having them function accordingly. Testing was done on our local environments prior to modularization to ensure that the Jupyter files for each program were converted into functional Python files. Once completed, the files were then modularized into separate plugin files & moved directly into each team's local PluMA framework. After successful tests from each team member on the team branch, the code is merged with the main Github branch.

SUMMARY

Significant progress has been made towards the creation of new ML plugins from 5 separate prediction model projects, with 3 projects being successfully modularized. Towards the beginning of the semester, every one of us was learning how to get acquainted with PluMA and determining a gameplan to tackle each prediction model. I had successfully achieved running the Flood-Prediction model on PluMA & modularized the model into two separate plugins.

REFERENCES & SCREENSHOTS



Research materials provided by Dr. Trevor Cickovski of Bioinformatics Research Group at Florida International University



**Engineering
& Computing**

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by Trevor Cickovski. We thank Trevor Cickovski for his assistance and mentorship that I received throughout the Capstone II project.

FLORIDA INTERNATIONAL UNIVERSITY