

Project: PluMA 2.0 Machine Learning Pipeline Refactor

Student: Michael Wei

Mentor: Dr. Trevor Cickovski, PhD, Florida International University, Bioinformatics Research Group (BioRG)

Instructor/Faculty: Dr. Masoud Sadjadi, PhD, Florida International University

PROBLEM

Bioinformaticians cannot reuse internal stages of existing pipelines often resulting in time reinventing the wheel that could be more optimally used to advance research. PluMA attempts to fix this problem by providing a pool where bioinformaticians can prototype algorithms and wrap them as plugins that can be reused by other users in other pipelines.

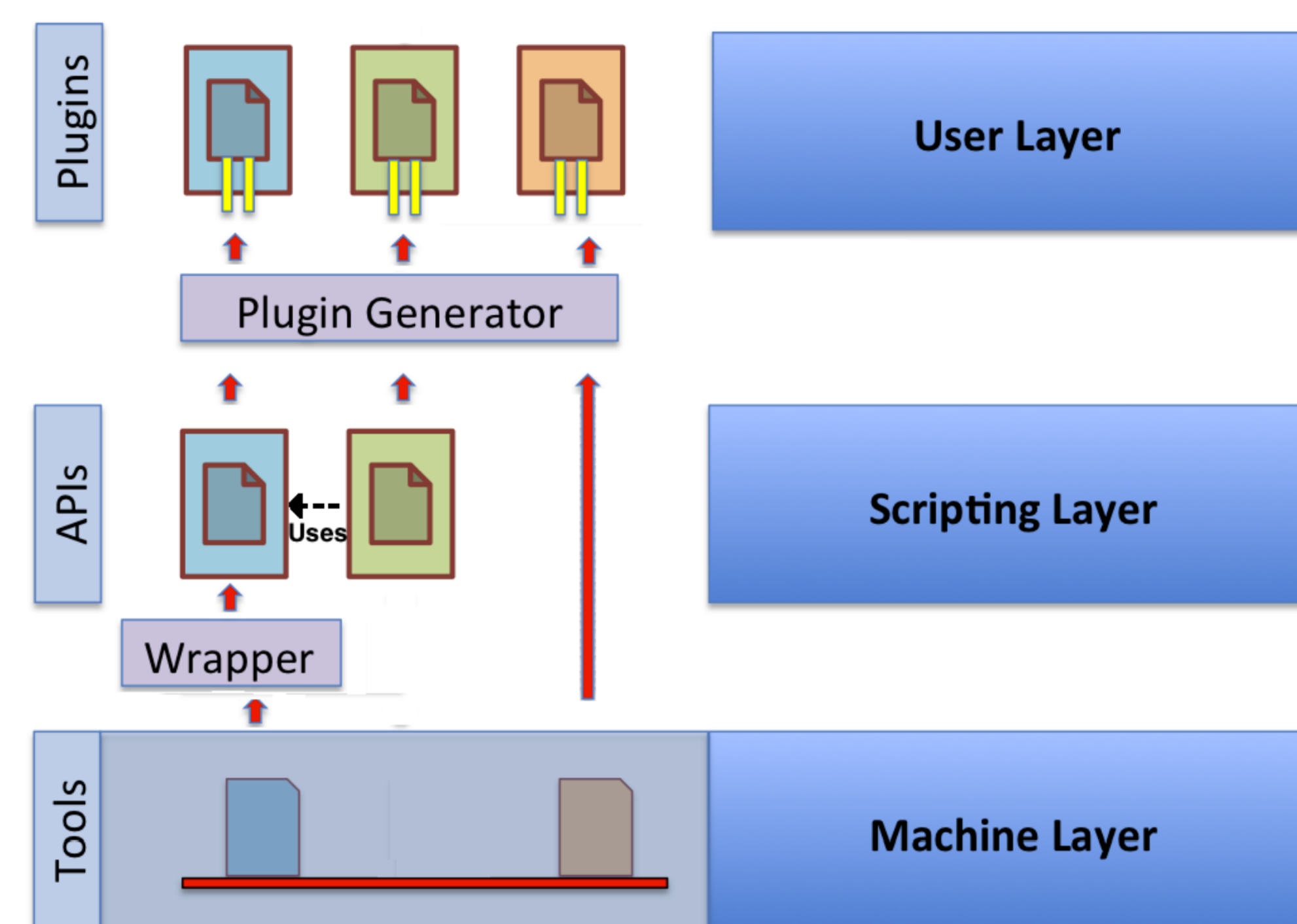
CURRENT SYSTEM

The inherently interdisciplinary bioinformatics landscape currently contains a great deal of pipelines, often implemented as single large scripts of functions.

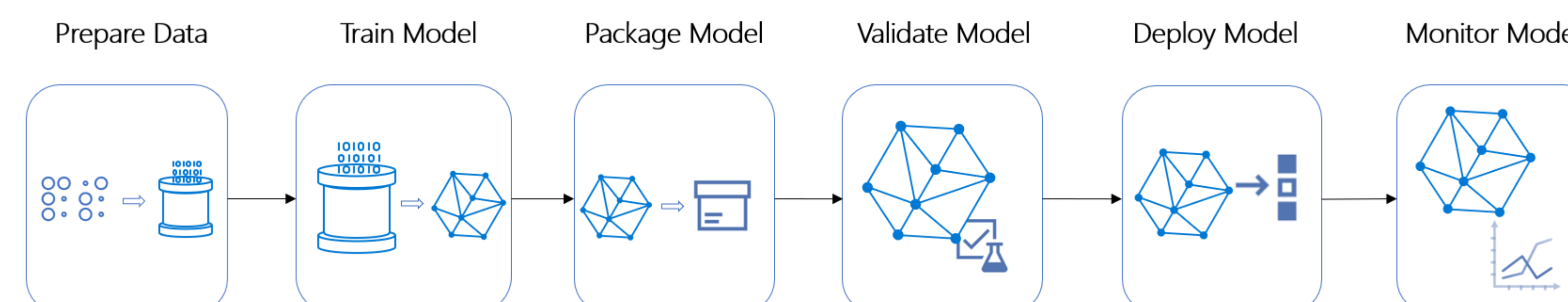
REQUIREMENTS

Properly refactor machine learning pipelines developed by collaborators with BioRG as a series of plugins to make compatible with PluMA

SYSTEM DESIGN



OBJECT DESIGN



IMPLEMENTATION



VERIFICATION

Files were tested during the processes of debugging and modularization to verify that they run as well as to check for other errors that may arise. Pipelines that were built at the end were tested to verify that they run.

SUMMARY

My contribution to the project involved debugging and modularizing files in the Jupyter folder. The process of debugging involved looking for module dependencies, file dependencies, and missing data. The process of modularization involved refactoring the code in the files and writing methods for the different sections of the code.

REFERENCES

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

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by Dr Trevor Cickovski. We thank Dr. Trevor Cickovski for his assistance and mentorship that we received throughout the senior design project.