

Knight Foundation School of Computing & Information Sciences

Summer 2022 Capstone 2

PluMA 2.0 Machine Learning Pipeline Refactor

Students: Gavin James

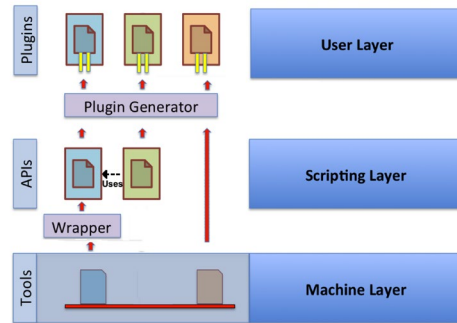
Mentor: Dr. Trevor Cickovski, PhD, Florida International University, Bioinformatics Research Group (BioRG)

Instructor/Faculty: Masoud Sadjadi, 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.

SYSTEM DESIGN



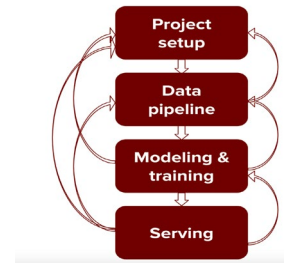
VERIFICATION

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.

CURRENT SYSTEM

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

OBJECT DESIGN



SUMMARY

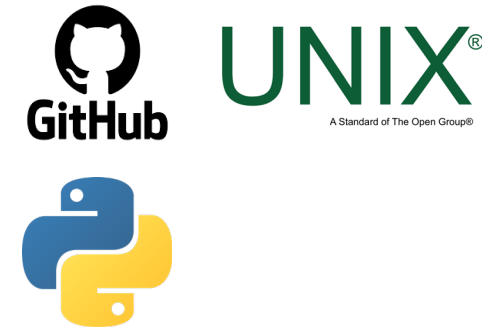
For this project

- I created outline of the lstm-flood-prediction folder within the FLOOD repository
- I created a Dependency List of the lstm-flood-prediction folder
- I created a ML template of the lstm-flood-prediction folder
- I edited and debugged the lstm-flood-prediction folder so it successfully ran in my local repository.
- I created a modularized architecture for the lstm-flood-prediction folder which allowed for reusable components of the code,

REQUIREMENTS

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

IMPLEMENTATION



REFERENCES

- Research and documentation materials provided by Dr. Trevor Cickovski of Bioinformatics Research Group at Florida International University
- “Design a Machine Learning System.” *Machine Learning Systems Design*, https://huyenchip.com/machine-learning-systems-design/assets/ml_project_flow.png

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by XXXXX (name of the project sponsor: federal, state or local government, or corporate partners, where applicable). We'll thank XXXX for his/her/their assistance and mentorship that I/we received throughout the senior design project.