Employing Machine Learning Models for CESM Timing Data





Why

Develop machine learning models that can be used to predict model performance for CESM from timing file data.

Implementation of the Project

Classification using acquired timing file data within Jupyter Notebooks to preserve workflow. Workflow can be repurposed for more complex scenarios.

Datasets, Preprocessing

- 3.5GHz Intel i5 vs. 2.5GHz Intel i7 dataset.
- **Principal Component Analysis** (PCA), Select K Best
- **Cheyenne component barriers** on versus component barriers off dataset (print statements commented).
- PCA, Select K Best, Recursive **Feature Elimination**
- **One Hot Encoding, Standard** Scaler

Key Results

	ē2
 3.5GHz Intel i5 vs. 2.5GHz Intel i7. Highest mean accuracy is 99.2% 	
from Decision Trees with Select K Best.	Gitl
 Lowest mean accuracy is 83.4% from Decision Trees with PCA. 	
Cheyenne component barriers on vs. component barriers off.	3
 Highest mean accuracy is 66.1% from SVMs with PCA. 	
 Lowest mean accuracy is 30.1% from random forests under Select K Best 	



Thomas Johnson III¹, Sheri Mickelson², Brian Dobbins², John Dennis² SIParCS Intern¹, SIParCS Mentors²

(Select K Best)





