

**University of Messina
MIFT Department
Bachelor Data Analysis**



UNIME Internship Project Proposal

<i>ID</i>	PTI_Distefano Salvatore_21/10/2023 17.00.46
<i>Date</i>	21/10/2023 17.00.46

Project Supervisor

<i>Surname</i>	Distefano
<i>Name</i>	Salvatore
<i>Department</i>	MIFT
<i>Laboratory</i>	HPCALab
<i>Email</i>	sdidtefano@unime.it
<i>Phone Number</i>	

Project Co-Supervisor (not compulsory)

<i>Surname</i>	Distefano
<i>Name</i>	Salvatore

<i>Job Position</i>	Professore
<i>Department</i>	MIFT
<i>Laboratory</i>	HPCALAB
<i>Email</i>	
<i>Phone Number</i>	

Internship Project Details

<i>Title</i>	Machine Learning WorkFlow MLWF
<p><i>Detailed Description:</i></p> <p>The final goal, the product of the machine learning process, is to create a model of the underlying system that is usually in part (gray box) or in total (black box) unknown, sometimes referred to as data-driven assessment. The mainstream approach mainly focuses on black box models but gray and white box ones can have several benefits. Exploiting the knowledge of a system, indeed, may reduce the problem complexity and consequently the amount of data required for training.</p> <p>One of the most interesting techniques adopted in the modeling theory to deal with complexity is to decompose or break down the original problem into subproblems through a top down approach. This comes as a natural fit when adopting model-driven white box assessment, since the internal structure of the system is usually known, broken down into subsystems/components and related interactions, even recursively/hierarchically. But it's quite harder in data-driven, black box approaches as ML.</p> <p>The research question this paper will try to answer is: is it possible to apply decomposition/partitioning to black box-ML models? Of course, the answer is yes, but how to without any information about the internal structure of a system?</p> <p>The main idea is to apply parallel computing and workflow theory, with specific regard to partitioning and decomposition models and techniques, to ML problems in a sort of ensemble learning fashion as done in https://doi.org/10.1016/j.eswa.2023.121772 (Fig. 5 - read the paper and the corresponding description to understand the approach).</p> <p>The main benefits of ML workflow decompositions can be assessed in terms of performance (greater accuracy, lower learning-training time, lower inference time) and in the size of the dataset required for training the ML models, breaking down the problem into smaller tasks reduces the data requirements.</p>	

<i>Duration (months – max 12)</i>	2
<i>Duration (hours)</i>	100
<i>Potential deadline</i>	
<i>Number of open position</i>	5

Internship Skills

<i>Required skills:</i> programming	
<i>Other skills</i>	ML, AI, microservices