Distributed Machine Learning
Instructor: Mohammadhadi Amini
Course Justification

Due to the increasing size of datasets, as well as privacy concerns of autonomous agents, the popularity of distributed machine learning technique has become widespread in recent years and the pervasiveness of this domain has been fueled by the means of advancing in machine learning and improvement of IoT infrastructures. In order to improve the prediction quality and render efficient machine learning solutions for complex real-life applications, an overwhelming amount of training data is needed. Though, a small amount of training data is sufficient for simple machine learning models, a larger and complicated model (e.g., artificial neural networks) expands exponentially corresponding to the number of parameters. To process such huge data streams in a centralized system requires computing infrastructures with a massive computation power. Alternatively, the machine learning workload can be distributed across multiple machines that turns the centralized systems into a distributed system that eliminates the issues with the machine learning model construction for complex networks. Further, it helps preserving the privacy of local entities by eliminating the need for sharing their data with a central server for the training purpose. It is crucial for computer science (CS) students, specifically those who are interested in real-time computing, distributed machine learning, efficient decision-making, and learning methods, to understand the theoretical aspects of the tools they deploy in research, being familiar with fundamental notions, and applying distributed machine learning tools and algorithms to solve relevant research problems.

There are courses on data science and machine learning at FIU, including Principles of Data Mining, Introduction to Data Science, Introduction to Machine Learning, Advanced Topics in Machine Learning, and Capstone in Data Science classes. Aforementioned courses comprehensively investigate the data science and machine learning algorithms and tools, including deep learning. It is imperative to develop a course that specifically provides the knowledge on how distributed machine learning algorithms can be leveraged for efficient learning in large-scale problems. The main goal of this course is to prepare students with the required concepts of distributed machine learning, and to understand the theoretical foundations of distributed machine learning, its relevant algorithms, and their potential applications. This graduate-level course (Distributed Machine Learning) will thoroughly describe distributed machine learning theory as a foundation for several areas in CS, including but not limited to efficient machine learning, secure computing, distributed computing, learning and data science, as well as engineering applications. Students will be equipped with strong theoretical foundations that can be applied to a wide spectrum of applications. This course will also strengthen their vision and understanding of the complexity behind various distributed machine learning techniques for efficient computing.

In order to expose student to real-world problems, practical use-cases from different application domains will be introduced during the semester. Students will be engaged in the learning process by solving hands-on problems during the course. This new course will provide CS students with a thorough understanding of theory behind several distributed machine learning tools and applications. Several senior positions in industry as well as national labs require such solid background. As the required preliminaries and mathematical concepts are covered in the class, upon instructor's permission, it also can benefit non-CS students in the College of Engineering and Computing who are interested in applying distributed machine learning techniques to their research for several computing and engineering applications.