



FLORIDA INTERNATIONAL UNIVERSITY
UNIVERSITY CURRICULUM COMMITTEE
Proposal for a New Course

DO NOT TYPE IN THIS BOX

Bulletin #: 6

Academic Year: 2020-21

1. School/College Engineering and Computing

Div./Dept. in Which Taught Knight Foundation School of Computing and Information Sciences

2. CIS 5 3
Alpha Prefix 1st Digit Last 3 Digits "C"-lec-lab "L"-Lab Cr. Hrs.

CIP Code (Leave this blank):

CAP 5617

3. Grading Method (select one): [X] Graded [ ] Pass/Fail

4a. Course Title Distributed Machine Learning

b. Abbreviated course Title (for computer class schedules, transcripts)

Distr. Machine Learning

LIMITED TO 25 Characters (including spaces)

5. Statewide Course Numbering Subject Matter Area CIS (Computer Science and Information Systems)

6. Catalog Description/Major Topics (not to exceed 200 characters including spaces)

College of Medicine and College of Law: Attach description not exceeding 1,000 characters including spaces.

Distributed Machine Learning for CS graduate students, including algorithms, efficient learning and computing tools, and real-world applications.

7. Attach detailed syllabus course outline and course justification on separate page(s).

8. Prerequisite(s): Introduction to Machine Learning (or instructor permission) Per dept., "CAP 4612 or CAP 5610 (or instructor permission)"

9. Corequisite(s): None

10. Objective(s) of Course:

Understand the concepts of distributed machine learning algorithms and tools, and their applications in efficient decision-making in real-world problems

11. Does this course duplicate/overlap other courses at FIU? [X] No [ ] Yes

If yes, please explain:

12. What other closely related department(s) have been consulted about this course?

13. Is this course used for the assessment of a program or a certificate (if yes, then send a notification to assessment@fiu.edu)? No Yes

PROPOSAL REQUESTED BY:

Faculty Contact Mohammadhadi Amini

(Type name)

moamini@fiu.edu

(Email address)

Chairperson (Dept./Div.) Jason Liu

(Type name)

Chairperson (Curr. Comm.) Wei-Chiang Lin

(Type name)

College/School Dean John Volakis

(Type name)

DocuSigned by:

Mohammadhadi Amini 13 / 2021

(Signature)

305-348-9936

(number)

3 15 / 2021

(Signature)

03 / 17 / 2021

(Signature)

Anthony McGoron 1 / 2021

(Signature)

Submit one original form. Attach one copy of the course justification and course syllabus, course description, objectives, major topics and textbooks.

## **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.

## School of Computing and Information Science

**Course Title:** Distributed Machine Learning

**Date:** 02/28/2021

**Course Number:** COT 5XXX

**Number of Credits:** 3

<b>Subject Area:</b> Intelligent Systems	<b>Subject Area Coordinator:</b> <b>email:</b>
<b>Catalog Description:</b> Distributed Machine Learning for CS graduate students, including algorithms, efficient learning and computing tools, and real-world applications.	
<b>Objective of Course:</b> Understand the concepts of distributed machine learning algorithms and tools, and their applications in efficient decision-making in real-world problems	
Textbook: [1] Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. Foundations of machine learning. MIT press, 2018. [2] Mitchell, Tom M. "Does machine learning really work?." AI magazine 18.3 (1997): 11-11.	
<b>References:</b> A collection of Relevant conference and journal papers.	
<b>Prerequisite Courses:</b> Introduction to Machine Learning (or instructor permission)	
<b>Corequisite Courses:</b> None	

**Type:** Elective

**Prerequisite Topics:**

- Calculus, Basic Programming (e.g., Python)
- Basic Machine Learning Concepts

**Course Outcomes:**

Students who successfully complete this course will be able to:

1. Explain the general concepts of distributed machine learning, algorithms, and its applications;
2. Formulate a distributed machine learning problem from scratch;
3. Apply the currently widely popular distributed machine learning algorithm, i.e., federated learning to real-world problems;
4. Identify and become familiar with the future research trends that can benefit from/contribute to distributed machine learning
5. Explain and discuss emerging real-world applications of distributed machine learning, e.g., Distributed machine learning for finance, improving resilience of critical infrastructures.

**Outline:**

Topic	Number of Lectures (Total: 37.5 hours = 15 weeks * 2 lectures/week * 1.25 hrs/lecture)	Outcome
<b>Course Overview and Preliminaries:</b> Course Overview and Introduction to Distributed Algorithms; Distributed Computing; Machine Learning concepts; Typical Supervised Learning Pipeline; Linear Algebra Review.	4	1
<b>Linear Regression and Distributed ML Principles:</b> Linear Regression; Distributed Machine Learning: Computation and Storage, (Stochastic) Gradient Descent; Communication Hierarchy; Communication Principles of Distributed Machine Learning; Scalable Machine Learning Tools.	7	1, 2
<b>Distributed Principal Component Analysis:</b> Exploratory Data Analysis; PCA Overview; PCA Assumptions and Solution; PCA Algorithm; PCA Derivation; Distributed PCA.	6	1, 2
<b>Comprehensive Overview of Federated Learning:</b> Federated Learning Overview; Categorization; Architecture; Lifecycle of Federated Learning; A Typical Federated Training Procedure; Handling Non-IID Data in Federated Learning; Optimization Algorithms for Federated Learning; Federated Learning Challenges; Resource-constrained Distributed Computing; Open Research Issues in Federated Learning Research Domain.	7	2, 3, 4
<b>Applications of Distributed Learning:</b> Distributed/Federated Learning for Computer Vision; Distributed Learning for IoT Applications; Distributed Learning for Finance; Federated Learning for Critical Infrastructure Resilience.	6	3, 4, 5