



DO NOT TYPE IN THIS BOX

Bulletin # : 3

Academic Year : 2018-19

FLORIDA INTERNATIONAL UNIVERSITY  
UNIVERSITY CURRICULUM COMMITTEE  
Proposal for a New Course

1. School/College Engineering and Computing

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

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

CIP Code (Leave this blank): \_\_\_\_\_

**CAP 5768C**

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

4a. Course Title Practical Data Science

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

Practical Data Science

LIMITED TO 25 Characters (including spaces)

5. Statewide Course Numbering Subject Matter Area CIS (Computing and Information Sciences)

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.

Topics will include: data collection and processing, data visualization and presentation, statistical model building using machine learning, and big data techniques for scaling these methods.

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

8. Prerequisite(s): M.S. or Ph.D. standing or permission of the instructor

9. Corequisite(s): None. M.S. & Ph.D. in CS students may not take this course for their degrees

10. Objective(s) of Course:

Understand the principles and approaches of big data analytics, including the linguistic phenomena to be modelled, the formal methods in use, and the potential end applications.

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

If yes, please explain: \_\_\_\_\_

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

Electrical Engineering, Economics, Business, Civil Engineering

PROPOSAL REQUESTED BY:

Faculty Contact Dong Chen

(Type name)

dochen@fiu.edu

(Email address)

(Signature)

(305) 348-7983

(Phone number)

11 / 28 / 2018

Chairperson (Dept./Div.) Ram Iyengar

(Type name)

(Signature)

11 / 28 / 2018

Chairperson (Curr. Comm.) Cesar Levy

(Type name)

(Signature)

12 / 04 / 2018

College/School Dean John Volakis

(Type name)

(Signature)

\_\_\_\_\_ / \_\_\_\_\_ / 2018

Submit one original form. Attach one copy of the course syllabus containing: course description, objectives, learning outcomes, major topics and textbooks.

an

2

10/10/10

L

## **CAP XXXX: Practical Data Science (Grad Course) Course Justification**

Data science is the study and practice of how we can extract insight and knowledge from large amounts of data. It is a burgeoning field, currently attracting substantial demand from both academia and industry.

There is no graduate course in data science at FIU that targets at students with non-CS majors. Such data-science skills are in high-demand and are essential in today's industry. This course provides a practical introduction to the "full stack" of data science analysis, including data collection and processing, data visualization and presentation, statistical model building using machine learning, and big data techniques for scaling these methods.

Specifically, topics to be covered include: introduction to data analysis techniques and associated computing concepts, foundations for data analysis, data visualization, parallel processing, metadata, ontologies, provenance, and data stewardship.

As the course name suggests, this course will highlight the practical aspects of data science, with a focus on implementing and making use of the above techniques. This new course will give non-CS major students data science skills along with sound computational mindset to better prepare them for today's marketplace.



## School of Computing and Information Science

**Course Title:** Practical Data Science

**Date:** 11/28/2018

**Course Number:** CAP XXXX

**Number of Credits:** 3

<b>Subject Area:</b> Data Science	<b>Subject Area Coordinator:</b> <b>email:</b>
<b>Catalog Description:</b> Practical Introduction to Data Science: computational thinking, data science, big data, data representation, data analysis, data visualization, meta data, and tools for data processing, analysis and visualization. For non-CS majors.	
<b>Textbook:</b> <i>Mining of Massive Datasets, 3<sup>rd</sup> Edition</i> . Jure Leskovec, Anand Rajaraman, and Jeff Ullman, 2018. ISBN-13: 978-1107077232	
<b>References:</b> None	
<b>Prerequisite Courses:</b> None	
<b>Corequisite Courses:</b> None	

**Type:** Elective

This course may not be used as a Core Course for the Computational Data Analysis track of the Master Degree program in Data Science.

**Prerequisite Topics:**

Pre-College Mathematics

**Course Outcomes:**

The objective is to develop computational thinking of the students. The implementation and coding will be kept to a minimum. The course will enable students to:

1. Describe the fundamental concepts, methods, and model for Data Science;
2. Learn fundamentals of Python programming for Data Science;
3. Acquire advanced computational thinking skills that will enable students to represent and reason about complex problems in the digital arena;
4. Understand different kinds of data in terms of their possibilities and limitations to approach complex problems cast in terms of the emerging field of Data Science;
5. Be comfortable with advanced data representation such as Graphs, Matrices and Linear Associated Arrays.
6. Become data science scholars through best practices in data documentation and dissemination.

**Outline:**

Topic	Number of Lecture Hours (Total: 37.5 hours)	Outcome
<b>Introduction to Data Science</b>	2.5	O1

<ul style="list-style-type: none"> <li>• Computational Thinking</li> <li>• Reasoning and Analysis</li> <li>• Data Science Challenges</li> <li>• The context of Data Science</li> </ul>		
<b>Data Representation</b> <ul style="list-style-type: none"> <li>• Time series data</li> <li>• Networked data</li> <li>• Big data</li> <li>• Spreadsheets, Databases</li> <li>• Matrices, Graphs</li> </ul>	2.5	O3
<b>Data Analysis Software</b> <ul style="list-style-type: none"> <li>• Programs for Data Analysis</li> <li>• Programming Languages</li> <li>• Algorithms versus Software</li> <li>• Data Structures for Data Sciences</li> </ul>	2.5	O3
<b>Data Analysis Workflows</b> <ul style="list-style-type: none"> <li>• Workflows for Data Analysis</li> <li>• Pre-processing, Post-processing Data</li> <li>• Workflow Inputs and Parameters</li> <li>• Exploring Data through Workflows</li> </ul>	2.5	O3
<b>Fundamentals of Python Programming for Data Science</b> <ul style="list-style-type: none"> <li>• Python Notebooks and Case Study</li> <li>• nd.Arrays, Series, Files and Maps</li> <li>• Python DataFrames and SQL</li> </ul>	2.5	O2, O3, O6
<b>Data Pre-processing</b> <ul style="list-style-type: none"> <li>• Data Cleaning</li> <li>• Quality Control</li> <li>• Data Integration</li> <li>• Feature Selection</li> <li>• Feature Construction</li> </ul>	2.5	O2, O3, O6
<b>Data Lifecycle</b> <ul style="list-style-type: none"> <li>• Data Collection and Storage</li> <li>• Data Extraction and Querying</li> <li>• Data Integration, Data Virtualization</li> </ul>	2.5	O3, O6
<b>Data Visualization Tools</b> <ul style="list-style-type: none"> <li>• Visualizations Quality and Types</li> <li>• Time Series Visualizations</li> <li>• Geospatial Visualizations</li> <li>• Multi-dimensional Spaces</li> <li>• Network Visualizations</li> </ul>	2.5	O6
<b>Data Analysis Tasks</b> <ul style="list-style-type: none"> <li>• Data Mining, and Machine Learning</li> </ul>	2.5	O1, O5

<ul style="list-style-type: none"> <li>• Supervised Learning (SVM, Random Forest, Decision Tree, Neural Networks)</li> <li>• Unsupervised Learning (Clustering)</li> <li>• Pattern Detection</li> <li>• Anomaly Detection</li> <li>• Probabilistic Graphical Models</li> <li>• Bayesian Networks, Causal Models</li> </ul>		
<b>Case Studies for Data Analysis</b> <ul style="list-style-type: none"> <li>• Text Data</li> <li>• Streaming Data</li> <li>• Time-series Data</li> <li>• Network Data</li> <li>• Image Data</li> </ul>	5	O5
<b>Tools for Big Data Analysis</b> <ul style="list-style-type: none"> <li>• Multi-core Computing</li> <li>• Distributed and Cluster Computing</li> <li>• Cloud and AWS Services</li> <li>• Parallel Programming Languages</li> <li>• MapReduce/Hadoop</li> </ul>	2.5	O3, O4, O6
<b>Semantic Metadata and Provenance</b> <ul style="list-style-type: none"> <li>• Metadata Collection and Processing</li> <li>• Metadata Search and Retrieval</li> <li>• Domain Metadata and Ontologies</li> <li>• Linked Open Data and Semantic Web</li> <li>• Provenance Models and Standards</li> </ul>	2.5	O3, O4
<b>Interdisciplinary Applications</b> <ul style="list-style-type: none"> <li>• Medical Science Diagnostic</li> <li>• Diagnosis or Prognostics</li> <li>• Drug Resistances, Infectious Diseases</li> <li>• Clinical Decision Support Systems</li> <li>• Molecular Diagnostics</li> </ul>	2.5	O5, O6
<b>Final Projects/Topics Presentation by Students</b>	2.5	O5, O6

