FLORIDA ATLANTIC UNIVERSITY™

Graduate Programs—NEW COURSE PROPOSAL[1]

UGPC APPROVAL _____
UFS APPROVAL _____
SCNS SUBMITTAL _____
CONFIRMED _____
BANNER POSTED _____
CATALOG _____

DEPARTMENT: DEPT. OF COMPUTER & ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

COLLEGE: COLLEGE OF ENGINEERING AND COMPUTER SCIENCE

RECOMMENDED COURSE IDENTIFICATION:

PREFIX ___CAP___ COURSE NUMBER ___6780___ LAB CODE (L or C) ____

(TO OBTAIN A COURSE NUMBER, CONTACT NMALDONADO@FAU.EDU)

COMPLETE COURSE TITLE: BIG DATA ANALYTICS WITH HADOOP

EFFECTIVE DATE
(first term course will be offered)

_____SPRING 2016_____

CREDITS[2]:
3

TEXTBOOK INFORMATION:
Data Mining: Practical Machine Learning Tools and Techniques, by I.H. Witten and E. Frank
Selected articles and papers

GRADING *(SELECT ONLY ONE GRADING OPTION)*: REGULAR __X___ SATISFACTORY/UNSATISFACTORY _____

COURSE DESCRIPTION, NO MORE THAN THREE LINES: THE STUDY OF TOPICS IN DATA MINING AND MACHINE LEARNING RELATING TO BIG DATA. BIG DATA CHALLENGES SUCH AS HIGH DIMENSIONALITY, CLASS IMBALANCE, QUALITY OF DATA, ETC. WILL BE EXAMINED AND ADDRESSED. HANDS-ON EXPERIENCE WITH BIG DATA ANALYSIS IN HADOOP USING A HIGH PERFORMANCE COMPUTING CLUSTER.

PREREQUISITES *:	COREQUISITES*:	REGISTRATION CONTROLS (MAJOR, COLLEGE, LEVEL)*:
GRADUATE STANDING OR PERMISSION OF INSTRUCTOR		GRADUATES IN COMPUTER ENGINEERING, COMPUTER SCIENCE, AND ELECTRICAL ENGINEERING.

* *PREREQUISITES, COREQUISITES AND REGISTRATION CONTROLS WILL BE ENFORCED FOR ALL COURSE SECTIONS.*

MINIMUM QUALIFICATIONS NEEDED TO TEACH THIS COURSE:
MEMBER OF THE GRADUATE FACULTY OF FAU AND HAS A TERMINAL DEGREE IN THE SUBJECT AREA (OR A CLOSELY RELATED FIELD)

Faculty contact, email and complete phone number:
Taghi Khoshgoftaar, khoshgof@fau.edu
561-297-3994

Please consult and list departments that might be affected by the new course and attach comments.[3]
ITOM (College of Business)
Mathematical Sciences (College of Science)

Approved by:	*Date:*
Department Chair: Hung Ender	9/9/15
College Curriculum Chair:	9/21/15
College Dean:	9/21/15
UGPC Chair: Wm. Mahmoud	10-7-2015
Graduate College Dean: Deborah Klein	
UFS President:	
Provost:	

1. **Syllabus** must be attached: see guidelines for requirements: *www.fau.edu/provost/files/course syllabus.2011.pdf*
2. Review **Provost Memorandum: Definition of a Credit Hour** *www.fau.edu/provost/files/Definition Credit Hour Memo 2012.pdf*
3. **Consent** from affected departments (attach if necessary)

Email this form and syllabus to UGPC@fau.edu **one week before** the University Graduate Programs Committee meeting so that materials may be viewed on the UGPC website prior to the meeting.

Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Course Syllabus

1. Course title/number, number of credit hours	
Big Data Analytics with Hadoop CAP 6780	3 credit hours
2. Course prerequisites, corequisites, and where the course fits in the program of study	
Prerequisites: Graduate standing or permission of instructor	
3. Course logistics	
Term: Spring 2016 This is a classroom lecture course with DL sections. *Class location and time*: Thursday 4:00 – 6:50 PM CM130	
4. Instructor contact information	
Instructor's name *Office address* *Office Hours* *Contact telephone number* *Email address*	Dr.Taghi M Khoshgoftaar, Professor Engineering East Bldg., Room 511 Tuesday and Thursday 11:00 AM – 2:00 PM 561-297-3994 khoshgof@fau.edu
5. TA contact information	
6. Course description	
The study of topics in data mining and machine learning relating to Big Data. Big Data challenges such as high dimensionality, class imbalance, quality of data, etc. will be examined and addressed. Hands-on experience with Big Data analysis in Hadoop using a high performance computing cluster.	
7. Course objectives/student learning outcomes/program outcomes	
Course objectives	Students will learn data mining and machine learning techniques for Big Data with Hadoop. Hands-on Big Data analysis using a high performance computing cluster. Case studies with an emphasis on real world applications will be presented.
BSCS program outcomes	
8. Course evaluation method	
Assignments (Homework, Programming, etc.) - 50% Term Project, Report – 35% Term Project, Presentation – 15%	The term project consists of a literature review of current state-of-the-art methods in advanced analytics with Big Data, or developing/advancing open source tools for machine learning with Big Data.
9. Course grading scale	
Grading Scale: 90 and above: "A", above 85 but below 90: "B+", 80-85: "B", above 75 but below 80: "C+", 70-75: "C", above 65 but below 70: "D+", 60-65: "D", above 55 but below 60: D-, 55 and below: "F."	

Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Course Syllabus

10. Policy on makeup tests, late work, and incompletes

Assignments are to be submitted on time, with possible point penalties for late submissions. In no case will an assignment be accepted after the graded papers for that assignment have been returned to the students. However, appropriate accommodations will be made for students having a valid medical excuse for being unable to work on an assignment during its two week period.

Unless there is solid evidence of medical or otherwise serious emergency situation incomplete grades will not be given.

11. Special course requirements

12. Classroom etiquette policy

University policy requires that in order to enhance and maintain a productive atmosphere for education, personal communication devices, such as cellular phones, are to be disabled in class sessions, and laptops are only to be used for note taking and related activities.
.

13. Disability policy statement

In compliance with the Americans with Disabilities Act (ADA), students who require special accommodations due to a disability to properly execute coursework must register with the Office for Students with Disabilities (OSD) located in Boca Raton campus, SU 133 (561) 297-3880 and follow all OSD procedures.

14. Honor code policy

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the university mission to provide a high quality education in which no student enjoys unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and place high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. See University Regulation 4.001 at www.fau.edu/regulations/chapter4/4.001_Code_of_Academic_Integrity.pdf

15. Required texts/reading

(1) Data Mining: Practical Machine Learning Tools and Techniques, by I.H. Witten and E. Frank
(2) Selected articles and papers are posted on the course web site. The list of selected papers is provided at the end.

16. Supplementary/recommended readings

17. Course topical outline, including dates for exams/quizzes, papers, completion of reading

Date:	Topic	Reading
Week 1	Introduction to Data Analytics	Ch 1-2
Week 2	Classification models Performance metrics	Ch 3-5

Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Course Syllabus

Week 3	Introduction to Big Data Hadoop/HDFS MapReduce	**Slides** Prepared teaching materials from many sources.
Week 4	Data sampling techniques for handling class imbalance Advanced classification models	Selected Articles
Week 5	H2O modeling tool for Big Data Using H2O to handle class imbalance with over/under-sampling **Homework 1 due**	**Slides** Prepared teaching materials from many sources.
Week 6	Feature selection techniques for handling high dimensionality	Selected Articles
Week 7	Spark big data processing engine and MLlib machine learning toolkit Feature engineering and feature selection with Spark/MLlib	**Slides** Prepared teaching materials from many sources.
Week 8	Ensemble learning **Homework 2 due**	Selected Articles
Week 9	Large-scale data processing with a Hadoop cluster Introduce cluster and how to use it	**Slides** Prepared teaching materials from many sources.
Week 10	Quality of data **Homework 3 due**	Selected Articles
Week 11	Quality of data Case Studies	Selected Articles
Week 12	Guest lecture presentation Students class (term project) presentations	
Week13	Students class (term project) presentations	
Week 14	Students class (term project) presentations **Homework 4 due**	
Week 15	**Term Project Due**	

List of Selected Papers

Logistic Regression Modeling of Software Quality
Taghi M. Khoshgoftaar, Edward B. Allen
International Journal of Reliability, Quality and Safety Engineering, Vol. 6, No. 4, 1999, pages 303-317

Experimental perspectives on learning from imbalanced data
Jason Van Hulse, Taghi M Khoshgoftaar, Amri Napolitano
Proceedings of the 24th international conference on Machine learning, 2007, pages 935-942

RUSBoost: A hybrid approach to alleviating class imbalance
Chris Seiffert, Taghi M Khoshgoftaar, Jason Van Hulse, Amri Napolitano
IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 40, No. 1, 2010, pages 185-197

Knowledge discovery from imbalanced and noisy data
Jason Van Hulse, Taghi M Khoshgoftaar
Journal of Data and Knowledge Engineering, Vol. 68, No. 12, 2009, pages 1513-1542

An empirical study of learning from imbalanced data using random forest
Taghi M Khoshgoftaar, Moiz Golawala, Jason Van Hulse
9th IEEE International Conference on Tools with Artificial Intelligence, 2007. ICTAI 2007, pages 310-317

Comparing boosting and bagging techniques with noisy and imbalanced data
Taghi M Khoshgoftaar, Jason Van Hulse, Amri Napolitano
IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 41, No. 3, 2011, pages 552-568

Note: This list may be updated in the future.

RE: Request from the CEECS Department

Tamara Dinev

To: Mihaela Cardei

Cc: Nurgun Erdol; Chiang-Sheng Huang; Caryn Conley

Tuesday, September 15, 2015 2:20 PM

Dear Dr. Cardei:

Regarding the 4 new course proposals below, I approve of their creation.

Regarding the Certificate in Big Data Analytics, per our conversation today with Dr. Erdol, rather than having two separate certificates in Data/Business Analytics, we agreed to create one certificate – in Big Data Analytics – with two tracks: Computer Science track and Business track. Students in each track with take 3 courses offered by the corresponding college, and one from the other college. Thus, a student in Computer Science track will take 3 CAP courses and 1 ISM course, and a student in College of Business will take 3 ISM courses and one CAP course.

Please contact Dr. Huang to coordinate how to amend our proposals toward this final version and fast track through the colleges so we can present our proposal at the upcoming University Council session.

Best Regards:
Tamara
==========

Tamara Dinev, Ph.D.
Department Chair and Professor
Department of Information Technology and Operations Management
College of Business
Florida Atlantic University
Boca Raton, Florida 33431
OFFICE: Fleming Hall, 219
TEL: (561) 297-3181
FAX: (561) 297-3043
e-mail: tdinev@fau.edu

From: Mihaela Cardei
Sent: Thursday, September 10, 2015 9:25 AM
To: Tamara Dinev <tdinev@fau.edu>
Cc: Nurgun Erdol <erdol@fau.edu>; Mihaela Cardei <mcardei@fau.edu>
Subject: Request from the CEECS Department

Dear Dr. Dinev

I am the chair of the Graduate Programs Committee in the Department of Computer & Electrical Engineering and Computer Science (CEECS) at FAU, and we are proposing a Certificate Program in Big Data Analytics.

Please find attached to this email the Certificate description and 4 new course proposals (CAP 6771, CAP 6780, CAP6688, and CAP6776) which are listed in the Certificate.

We would need you approval that ITOM Department supports the Certificate in Big Data Analytics and the 4 new courses.

Could you please review the material and email me your approval decision?

Thank you,

Mihaela Cardei, PhD
Professor
Computer & Electrical Engineering and Computer Science Department
College of Engineering and Computer Science
Florida Atlantic University
http://www.cse.fau.edu/~mihaela

Re: Request for approval - Big Data Analytics Certificate & new courses

Rainer Steinwandt [srainer@math.fau.edu]

To: Mihaela Cardei

Wednesday, September 16, 2015 8:24 PM

Dear Mihaela,

Thank you for your email. The proposed certificate program and the associated courses of the CEECS Department and ITOM look very fine to me. For the Department of Mathematical Sciences, I support this certificate program and the associated courses and hope that this program will be a great success.

Kind regards,
Rainer

----- Original Message -----
From: "Mihaela Cardei" <mcardei@fau.edu>
To: "Rainer Steinwandt" <srainer@math.fau.edu>
Cc: "Nurgun Erdol" <erdol@fau.edu>, "Tamara Dinev" <tdinev@fau.edu>, "Chiang-Sheng Huang" <dhuang@fau.edu>, "Mihaela Cardei" <mcardei@fau.edu>
Sent: Wednesday, September 16, 2015 7:26:41 PM
Subject: Request for approval - Big Data Analytics Certificate & new courses

Dear Dr. Steinwandt,

The Department of Computer & Electrical Engineering and Computer Science (CEECS) and the Department of Information Technology and Operations Management (ITOM) at FAU are proposing a joint Certificate Program in Big Data Analytics, with two tracks: Computer Science and Business.

In addition, CEECS Department is proposing 4 new course proposals (CAP 6771, CAP 6780, CAP6688, and CAP6776) and ITOM is proposing 3 new course proposals (ISM6422, ISM6119, ISM6058).

Please find attached to this email the Certificate and new course proposal documents.

We would need your approval that the Department of Mathematical Sciences supports the joint Certificate in Big Data Analytics and the new course proposals.

Could you please review the material and email me your approval decision?

Thank you,

Mihaela Cardei, PhD
Professor
Computer & Electrical Engineering and Computer Science Department
College of Engineering and Computer Science
Florida Atlantic University
http://www.cse.fau.edu/~mihaela