



**Coversheet - Proposal for New and Revised Courses**  
(Use for non-Pathways courses)

For CLE/Pathways courses, form can be found here: <https://www.pathways.prov.vt.edu/proposal-forms.html>

General Information			
Proposal Date:	2/10/21	Department:	Computer Science
Course Designator and Number (Cross-listed Course Designator and Number):	CS 5644		
Title of Course:	Machine Learning with Big Data	Credit Hours:	3
Course Transcript (ADP) Title (30 Characters & Spaces Maximum):	Machine Learning with Big Data		
Instructor and/or Departmental Contact:	Trey Mayo & Nicole Akers		
Contact Phone:	(540) 231-0780	Contact E-mail:	treymayo@vt.edu
Please refer to Office of University Registrar for guidelines and policy requirements: <a href="https://registrar.vt.edu/governance.html">https://registrar.vt.edu/governance.html</a>			

Please count this course toward the following Scorecard Metrics areas:

- Study Abroad       Service Learning       Experiential       Undergraduate Research

Scorecard Metrics Definitions can be found here: <https://registrar.vt.edu/faculty-toolbox/scorecard-metrics.html>

Please insert an X if this course should count toward First Year Experience:

- First Year Experience (FYE) Include approval letter from FYE Director. More information can be found here: <http://www.fye.vt.edu>

Select ONE of the following boxes	
<input type="checkbox"/> New Course	<input checked="" type="checkbox"/> *Revised Course (Revision > 20% _____ Revision < 20% <u>X</u> _____)

For CLE/Pathways courses, form can be found here: <https://www.pathways.prov.vt.edu/proposal-forms.html>

\*Please include a summary of course revisions to the Justification section of proposal

<b>A:</b>	Attach statement from Dean or Departmental Representative as to whether teaching this course will require or generate the need for additional departmental resources.		
<b>B:</b>	Attach appropriate letters of support (e.g., prerequisite, corequisite, or cross-list memo) from affected departments and/or colleges.		
<b>C:</b>	Effective Semester:	Summer 2021	
<b>D:</b>	Change in Title From:		
	To:		
<b>E:</b>	Change in Transcript Title (ADP) From:	To:	
<b>F:</b>	Change in Credit Hours From:	To:	
<b>G:</b>	Change in Lecture and/or Lab Hours From:	To:	
<b>H:</b>	Course Number(s) and Title(s) to be deleted from the Catalog with <u>APPROVAL</u> :		

Approval Signatures			
Department Representative		Date	2/15/2021
College Curriculum Committee Rep		Date	
College Dean or Designee		Date	3/15/21

## Course Information

### Catalog Description

Basic principles and techniques for big data analytics, including methods for storing, searching, retrieving, and processing large datasets; introduction to basic machine learning libraries for analyzing large datasets; data visualization; case studies with real-world datasets. Not for graduate credit for MS and PhD degrees in Computer Science and Applications (CSA); MEng degrees in CSA allowed to receive credit. Pre: 5044 (3H, 3C).

### Learning Objectives

Having successfully completed this course, the student will be able to:

- 1.
- 2.

### Justification

The changes to the course restriction are to differentiate that the MS and PhD degrees in CSA cannot receive credit, but the MEng can receive credit. Given the practitioner nature of the MEng degree, students in this particular degree program may find it appropriate to take this course to enhance the requirements for their intended career path.

### Prerequisites and Corequisites

Pre: 5044 Object-Oriented Pgmning Java



Texts and Special Teaching Aids

Please identify specific examples and whether these are Required or Recommended. If no required text, provide justification and include examples of "Recommended" materials that will be used.

Topic Syllabus

<u>Topic</u>	<u>Percent of Course</u>
Example Topic 1	
Example Topic 2	
Subtopic, as applicable when topics are >20%	
Subtopic	
	Total: 100%

Old (Current) Topic Syllabus

N/A for new courses. For an existing course that is being revised with a new course number, including the syllabus can be helpful. Copy-and-paste the topic syllabus from the proposal on file.

**From:** [Trey Mayo](#)  
**To:** [Akers, Nicole](#)  
**Subject:** CS 5644 Revision  
**Date:** Monday, March 15, 2021 4:15:37 PM

---

Nicole,

I am writing to let you know that the proposed changes to CS 5644 will not generate the need for additional resources.

Please let me know if you have any questions.

Thank you.

Trey Mayo

=====

**Trey Mayo, M.S.Ed.**

Director of Graduate Programs  
Department of Computer Science  
Virginia Tech  
[treymayo@vt.edu](mailto:treymayo@vt.edu)  
<http://cs.vt.edu>

“Too bad all the people who know how to run the country are busy driving cabs and cutting hair.” - George Burns

“Education is not preparation for life; education is life itself.” - John Dewey

“It’s not the years in your life that count. It’s the life in your years.” - Abraham Lincoln

CM-2868

### COVER SHEET - NEW and REVISED COURSES

Commission on Undergraduate Studies and Policies/ Commission on Graduate Studies and Policies/ University Curriculum Committee For Liberal Education  
Effective September 2010

**ENTERED**  
5/13/16  
SOS

•SEE I - VIII for Basic Course Proposal Guidelines•  
•SEE APPENDIX FOR NOTES, EXPLANATIONS AND ADDITIONAL GUIDELINES•  
•PRINT CLEARLY, TYPE or COMPLETE ELECTRONICALLY•

**APPROVED**  
GCC 4-28-16  
CGSP 5-4-16

**PROPOSAL DATE:** Feb. 8, 2016      **15-DAY REVIEW END DATE:**

**DEPARTMENT:** Department of Computer Science

**COURSE DESIGNATOR AND NUMBER:** CS 5644

**TITLE OF COURSE:** Machine Learning with Big Data

**TRANSCRIPT (ADP) TITLE (MAX-30 Characters):** Machine Learning with Big Data

**INSTRUCTOR and/or DEPARTMENTAL CONTACT:** Naren Ramakrishnan      **CONTACT MAILCODE:** 0379

**CONTACT PHONE:** 571-858-3331      **CONTACT E-MAIL:** naren@cs.vt.edu

Please count this course toward the following scorecard metrics area:  
 Study Abroad     Service Learning     Experiential     Undergraduate Research

CHECK ONLY ONE OF THE FOLLOWING BOXES

- NEW COURSE**       **REVISED COURSE** [Revision>20%\_\_\_\_ Revision<20%\_\_\_\_ ]
- NEW COURSE & INCLUSION IN THE CLE** [Area\_\_\_\_ ]       **OTHER:** \_\_\_\_\_  
Include Attachment, if Needed
- REVISED COURSE FOR INCLUSION IN THE CLE OR CLE AREA CHANGE**

•Courses routed directly to the University Curriculum Committee For Liberal Education MUST be endorsed by the appropriate Department Head or Dean.  
•The Chair of the University Curriculum Committee For Liberal Education shall inform the appropriate college curriculum committee of all courses under review by the University Curriculum Committee For Liberal Education.

•A **Attach Statement from Dean or Departmental Representative as to whether Teaching this Course will Require or Generate the Need for Additional Departmental Resources.**

•B **Attach Appropriate Letters of Support from Affected Departments and/or Colleges.**

•C **Effective Semester:** Fall 2016

•D **Change in Title From:**  
\_\_\_\_\_  
**To:** \_\_\_\_\_

•E **Change in Lecture and/or Lab Hours**      **From:** \_\_\_\_\_      **To:** \_\_\_\_\_

•F **Change in Credit Hours**      **From:** \_\_\_\_\_      **To:** \_\_\_\_\_

•G **Percentage of Revision from Current Syllabus:** \_\_\_\_\_      **Revision Summary:** \_\_\_\_\_

•H **Course Number(s) and Title(s) to be Deleted from the Catalogue with APPROVAL of course:**  
\_\_\_\_\_

**APPROVAL SIGNATURES**

Department Representative: Naren Ramakrishnan      Date: April 6, 2016  
College Curriculum Committee

Department Representative \_\_\_\_\_ Date: \_\_\_\_\_

College Curriculum Committee

Representative \_\_\_\_\_ Date: \_\_\_\_\_

College Dean \_\_\_\_\_ Date: \_\_\_\_\_

# COVER SHEET - NEW and REVISED COURSES

Commission on Undergraduate Studies and Policies/ Commission on Graduate Studies and Policies/ University Curriculum Committee For Liberal Education  
Effective September 2010

- SEE I - VIII for Basic Course Proposal Guidelines•
- SEE APPENDIX FOR NOTES, EXPLANATIONS AND ADDITIONAL GUIDELINES•
- PRINT CLEARLY, TYPE or COMPLETE ELECTRONICALLY•

PROPOSAL DATE: March 14, 2016 15-DAY REVIEW END DATE:

DEPARTMENT: Computer Science

COURSE DESIGNATOR AND NUMBER: CS 5644

TITLE OF COURSE: Machine Learning with Big Data

TRANSCRIPT (ADP) TITLE (MAX-30 Characters):  
Machine Learning with Big Data

INSTRUCTOR and/or DEPARTMENTAL CONTACT: Naren Ramakrishnan CONTACT MAILCODE: 0379

CONTACT PHONE: 571-858-3331 CONTACT E-MAIL: naren@cs.vt.edu

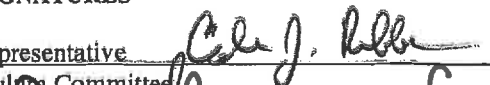
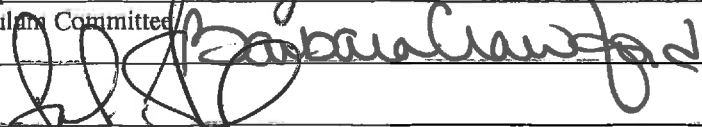

Please count this course toward the following scorecard metrics area:  
 Study Abroad  Service Learning  Experiential  Undergraduate Research

CHECK ONLY ONE OF THE FOLLOWING BOXES

- NEW COURSE  REVISED COURSE [Revision>20%\_\_\_\_\_ Revision<20%\_\_\_\_\_ ]
- NEW COURSE & INCLUSION IN THE CLE [Area\_\_\_\_\_ ]  OTHER: \_\_\_\_\_  
Include Attachment, if Needed
- REVISED COURSE FOR INCLUSION IN THE CLE OR CLE AREA CHANGE

•Courses routed directly to the University Curriculum Committee For Liberal Education MUST be endorsed by the appropriate Department Head or Dean.  
•The Chair of the University Curriculum Committee For Liberal Education shall inform the appropriate college curriculum committee of all courses under review by the University Curriculum Committee For Liberal Education.

- A Attach Statement from Dean or Departmental Representative as to whether Teaching this Course will Require or Generate the Need for Additional Departmental Resources.
- B Attach Appropriate Letters of Support from Affected Departments and/or Colleges.
- C Effective Semester: \_\_\_\_\_
- D Change in Title From: \_\_\_\_\_  
To: \_\_\_\_\_
- E Change in Lecture and/or Lab Hours From: \_\_\_\_\_ To: \_\_\_\_\_
- F Change in Credit Hours From: \_\_\_\_\_ To: \_\_\_\_\_
- G Percentage of Revision from Current Syllabus: \_\_\_\_\_ Revision Summary: \_\_\_\_\_
- H Course Number(s) and Title(s) to be Deleted from the Catalogue with APPROVAL of course: \_\_\_\_\_

APPROVAL SIGNATURES  
Department Representative:  Date: 3/14/16  
College Curriculum Committee Representative:  Date: 3/21/16  
College Dean:  Date: 3/22/16

# Machine Learning with Big Data

## CS 5644

### I. Catalog Description

Basic principles and techniques for big data analytics, including methods for storing, searching, retrieving, and processing large datasets; introduction to basic machine learning libraries for analyzing large datasets; data visualization; case studies with real-world datasets. Not for graduate credit for degrees in Computer Science and Applications (CSA) Pre: 5044. (3H, 3C)

**Course Number:** 5644

**ADP TITLE:** Machine Learning with Big Data

### II. Learning Objectives

Having successfully completed this course, students will be able to:

- Describe challenges associated with handling large datasets.
- Store, search, retrieve, and process large datasets using modern computing methodologies and data processing pipelines.
- Apply widely used machine learning tools and libraries to implement basic machine learning models (regression, classification, clustering, time-series analysis) and apply them to large datasets.
- Comparatively analyze machine learning methods and evaluate their performance.
- Communicate patterns discovered by the machine learning models from large datasets using basic data visualization techniques implemented using modern object-oriented libraries.

### III. Justification

The ubiquity of social media, sensor networks, and technological advances in data collection have resulted in massive datasets that demand methods for analysis and visualization. CS 5644 introduces the practical challenges associated with storing, retrieving, accessing and processing large datasets, machine learning methods for inferring patterns from data, and their subsequent interpretation using data visualizations. The course requires students to implement basic data analytic pipelines using machine learning libraries, such as Weka (Java) and scikit-learn (Python).

This course is designed for students pursuing a graduate certificate or degree as part of the Master of Information Technology program.

Graduate credit is required for this course. Students will apply advanced, extensive, and in-depth knowledge that builds on undergraduate learning with analysis of large datasets. Students will develop the ability to analyze and investigate massive datasets independently using machine learning techniques, and to deepen their scholarly development in empirical research into large-scale dataset analysis.

### IV. Prerequisites and Corequisites

Pre: 5044.



## V. Texts and Special Teaching Aids

### Required Textbooks:

Chodorow, Kristina, MONGODB: THE DEFINITIVE GUIDE. O'Reilly Publishers, 2013, 409

Witten, Ian H. and Eibe Frank. DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES. Morgan Kaufmann, 2005, 524

Garreta, Raul and Guillermo Moncecchi. LEARNING SCIKIT-LEARN: MACHING LEARNING IN PYTHON. Packt Publishing Ltd, 2013, 103

## VI. Syllabus

	<b>Percent of Course</b>
1. Introduction to the Python programming language	10
a. In-built modules in Python	
b. Implementing basic programs in Python	
c. Object-oriented programming in Python	
2. Storing, searching, accessing and processing large datasets	20
a. SQL, MySQL	
b. MongoDB, Redis	
c. High-level programming languages (Java, Python)	
3. Implementing machine learning models	20
a. Libraries for machine learning (Weka, Scikit-learn)	
b. Regression models	
c. Classification (Decision trees and logistic regression)	
d. Clustering (k-means)	
e. Time-series analysis and forecasting	
4. Data Visualization	25
a. Libraries for data visualization (Python, Java)	
b. Scalar visualization	
c. Vector visualization	
5. Specific case studies on real-world datasets	25
a. Social media analysis	
b. Electronic health records	
c. Sensor networks	
<b>TOTAL</b>	<b>100</b>

**TO:** Course Approval Committees  
**FROM:** Calvin Ribbens *Cal Ribbens*  
Department Head  
**RE:** CS 5644 course proposal  
**DATE:** March 14, 2016

The Department of Computer Science is requesting approval of a new course, CS 5644, "Machine Learning with Big Data." This course will be offered as part of the Masters of Information Technology (MIT) program.

No additional resources will be required in order to offer this course.