UNIVERSITY REGISTRAR
VIRGINIA TECH

CM-7467
Approved
GCC: 04.14.22
CGPSP: 04.20.22

ENTERED
5.12.22 ~RP

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:	2/1/22	*Department:*	Computer Science
Course Designator and Number (Cross-listed Course Designator and Number):	CS 5045-5046		
Title of Course:	Computation for the Data Sciences	*Credit Hours:*	3
Course Transcript (ADP) Title (30 Characters & Spaces Maximum):	Computation for Data Sciences		
Instructor and/or Departmental Contact:	Trey Mayo		
Contact Phone:	(540) 231-0780	*Contact E-mail:*	treymayo@vt.edu
Please refer to Office of University Registrar for guidelines and policy requirements: https://registrar.vt.edu/governance.html			

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
☐ New Course ☒ *Revised Course (*Revision > 20% ___X____ Revision < 20% _____*)*

*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

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*** (e.g., prerequisite, corequisite, or cross-list memo) from affected departments and/or colleges.			
C:	***Effective Semester:***	Fall 2022		
D:	***Change in Title From:***			
	To:			
E:	***Change in Transcript Title (ADP) From:***		***To:***	
F:	***Change in Credit Hours From:***		***To:***	
G:	***Change in Lecture and/or Lab Hours From:***		***To:***	
H:	***Course Number(s) and Title(s) to be deleted from the Catalog with APPROVAL:***			

Approval Signatures			
Department Representative		*Date*	1/26/22
College Curriculum Committee Rep		*Date*	2/12/22
College Dean or Designee		*Date*	2/12/22

Course Information

Catalog Description

Covers fundamentals of computer science and background in data sciences needed by graduate students without a computer science background. 5045: Programming language syntax and semantics for data science; abstraction and object-oriented programming; data structures; databases; visualization; ethics and data manipulation. 5046: Software engineering; data preprocessing; and machine learning. Pre: Graduate standing for 5045; 5045 for 5046. (3H, 3C)

Learning Objectives

CS 5045:
Having successfully completed this course, the student will be able to:

1. Design and critique useful programs of moderate size in the studied programming language.
2. Conceptualize data models and structures in data science programming languages.
3. Import raw data into data science programming platform.
4. Create a visualization of multiple and varying data objects.
5. Clean, manipulate, and transform imported data.
6. Provide a clean and prepared data set given a specified objective.
7. Assess the ethics of data collection/usage and the impacts to privacy.

CS 5046:
Having successfully completed these courses, the student will be able to:

1. Design software testing strategies.
2. Utilize software libraries in a typical programming language.
3. Preprocess large data sets.
4. Utilize and organize discipline-specific data sets to build machine learning models.

Justification

These courses provide essential data science and computer science material required for Masters and Ph.D. students in many departments. They provide for graduate students, otherwise lacking a computer science background, a rapid grounding in the computer science and data science fundamentals most essential for the pursuit of career paths that employ the analysis and storage of large modem data sets. This course is needed to support the Data and Decisions Destination Area at Virginia Tech and the Applied Data Science MS.

This revision revises the first course to focus more on data structures and the manipulation/visualization of data instead of focusing on algorithms. Data Science requires the use of computational languages to ingest and prepare data for data models and analysis. This course educates students on methods to prepare the data for those models/analysis. Typical estimates are that 80% of data science are tasks related to data preparation. CS 5046 remains unchanged with no revisions.

The course is taught at the 5000 level because the course focuses on doing advanced study and research in one of the many disciplines that employ data science. Students will gain programming, algorithm analysis, software engineering, and data analysis skills as a blended totality only available within the context of graduate study of such a discipline.

Prerequisites and Corequisites

Pre: Graduate standing for 5045. CS 5045 Computation for the Data Sciences for 5046.

Texts and Special Teaching Aids

Required for CS 5045:
McKinney, W. (2017). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* O'Reilly Media, Inc., pp. 550.

Miller, B.N., Ranum, D.L., & Anderson, J. (2019). *Python programming in context.* Jones & Bartlett Learning. pp. 498

Required for CS 5046:
Raschka, S., Julian, D., & Hearty, J. (2016). *Python: Deeper insights into machine learning.* Packt Publishing. pp. 916

VanderPlas, J. (2017). *Python data science handbook.* O'Reilly. pp. 530

Topic Syllabus

For CS 5045:

Topic	Percent of Course
An overview of programming and data representation	10%
Python Fundamentals a. Data types b. Control structures and functions c. Object-oriented programming a. Objects b. Classes c. Parameter Passing d. Inheritance d. File handling	35%
Data Structures a. Lists b. Dictionaries c. Tuples d. Relational	20%
Exploratory Data Analysis/Visualization a. Pandas b. Data Tables c. GroupBy d. Summary Functions (sum, mean, etc.) e. Plotting	25%
Data Manipulation a. Reshaping b. Join/Merge	10%

Total: 100%

For CS 5046:

1. Software engineering; development and testing - 20%
2. Python software libraries - 15%
3. Data preprocessing - 10%
4. Clustering techniques - 10%
5. Regression - 10%
6. Support vector machines - 10%
7. Decision trees - 10%
8. Artificial neural networks, deep learning - 15%

Total: 100%

Old (Current) Topic Syllabus

Topic	Percent of Course
An overview of algorithms, data representation, and programming	10%
Python Fundamentals a. Data types b. Control structures and functions c. Object-oriented programming d. File handling	40%
Problems, algorithms, asymptotics a. Divide and conquer, sorting b. Greedy algorithms c. Dynamic programming, sequence alignment d. Graph algorithms e. P, NP, and NP-completeness	40%
Databases	10%

Total: 100%

For CS 5046:

1. Software engineering; development and testing - 20%
2. Python software libraries - 15%
3. Data preprocessing - 10%
4. Clustering techniques - 10%
5. Regression - 10%
6. Support vector machines - 10%
7. Decision trees - 10%
8. Artificial neural networks, deep learning - 15%

Total: 100%

Virginia Tech
VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

Department of Computer Science
College of Engineering

2000A Torgersen Hall, Blacksburg, Virginia 24061-0106

(540) 231-4354
Email: shaffer@vt.edu
WWW: http://people.cs.vt.edu/~shaffer

March 12, 2022

TO: Course Approval Committees

FROM: Cliff Shaffer
Associate Department Head for Graduate Studies

RE: CS 5045

The Department of Computer Science is requesting approval for a course revision proposal for CS 5045 “Computation for the Data Sciences”.

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