RARITAN VALLEY COMMUNITY COLLEGE
ACADEMIC COURSE OUTLINE

CSIT 107 INTRODUCTION TO DATA SCIENCE

I. Basic Course Information

A. Course Number and Title: CSIT 107 Introduction to Data Science

B. New or Modified Course: **New**

C. Date of Proposal: Semester: **Fall** Year: **2024**

D. Effective Term: Fall 2025

E. Sponsoring Department: **Mathematics and Computer Science**

F. Semester Credit Hours**: 4**

G. Weekly Contact Hours**: 5**
Lecture**: 3**
Laboratory**: 2**
Out of class student work per week: **7**

H. ☒ Prerequisite (s): **MATH 030 Intermediate Algebra**
☐ Corequisite (s):

I. Additional Fees**:** None

II. Catalog Description

Prerequisite: MATH 030 – Intermediate Algebra
Introduction to Data Science will provide students with data literacy skills needed to understand the role of data in society as well as the tools and techniques to retrieve, tidy, clean and visualize data in preparation for statistical analysis. R and the R Studio environment will be used to work on assignments and projects that require both programming skill and statistical reasoning. A unique feature of this course is the use of R to demonstrate statistical concepts using simulations. No prior knowledge of this is required. Prediction and classification in machine learning models will be introduced as well as ethical issues surrounding the use of artificial intelligence and machine learning.

III. Statement of Course Need

A. Data Science has been and still is an up and coming area for all disciplines. This course will teach students the basics of Data Science using R programming and can be applied to many disciplines.

B. This course has a lab component. The lab is essential for providing students with applied hands-on experience with R and R Studio. Students are required to use the software in the computer labs to complete their assignments.

C. This course generally transfers as a Data Science program requirement dependent on the transfer institution.

IV. Place of Course in College Curriculum

A. Free Elective
B. This course serves as a General Education course in Mathematics and Technological Competency, pending state approval.
C. This course serves as a Programming Elective in the Computer and Programming Electives List.
D. This course meets a program requirement for the Data Science Option in the Computer Science AS degree.
E. To see course transferability: a) for New Jersey schools, go to the NJ Transfer website, www.njtransfer.org ; b) for all other colleges and universities, go to the individual websites.

V. Outline of Course Content

A. Introduction to R Programming
B. Data Visualization with ggplot2
C. Exploratory Data Analysis
D. Feature Engineering
E. Probability and Statistics Concepts with R
    a. Generating random data
    b. Normal Distribution
    c. Sampling
    d. Law of large numbers
    e. Sampling distributions
F. Statistical Concepts with R
    a. Confidence intervals
    b. Confidence level
    c. Hypothesis test
    d. P-value
    e. Significance level
    f. Errors
G. Intro to Machine Learning
    a. Linear Regression

b. K-Means Clustering
c. Learning with Biased Data

VI. A. Course Learning Outcomes:

At the completion of the course, students will be able to:

1. Import, modify and reshape data (including messy or incomplete data) to formats suitable for data analysis. (GE -4)
2. Produce and interpret data visualizations to describe and explore large datasets. (GE -2,4*)
3. Solve data science problems using the R programming language. (GE -4*)
4. Perform statistical inference in R and interpret related values, e.g. p-value, significance level, confidence level, margin of error. (GE -2,4)
5. Implement supervised or unsupervised machine learning algorithms. (GE -4)
6. Discuss ethical issues in Machine Learning and Artificial Intelligence. (GE -4, ER)

*critical thinking

B. Assessment Instruments

1. Projects
2. Exams
3. Labs
4. Coding Challenges

VII. Grade Determinants

A. Projects
B. Exams
C. Labs
D. Coding Challenges

Given the goals and outcomes described above, LIST the primary formats, modes, and methods for teaching and learning that may be used in the course:

A. lecture/discussion
B. small-group work
C. computer-assisted instruction
D. laboratory
E. student oral presentations
F. student collaboration

VIII. Texts and Materials

A. suggested textbooks

Thulin, M. (2024). Modern Statistics with R. Second edition.
Free for students https://modernstatisticswithr .com

R for Data Science Authors: Garrett Grolemund and Hadley Wickham
Free for students: https://r4ds.had.co.nz/

Please Note: The course outline is intended only as a guide to course content and resources. Do not purchase textbooks based on this outline. The RVCC Bookstore is the sole resource for the most up-to-date information about textbooks.

IX. Resources

A. Computer lab for classroom instruction and exercises.
B. R and R Studio

X. Check One: ☐Honors Course ☐Honors Options ☒ N/A