

Course Outline

General Information

COURSE ID (CB01A AND CB01B)

CIS 9

COURSE TITLE (CB02)

Introduction to Data Science

COURSE CREDIT STATUS

Credit - Degree Applicable

EFFECTIVE TERM

Fall 2021

COURSE DESCRIPTION

This course is an introduction to data science, which covers data analytics and machine learning. Topics covered include data gathering and data wrangling, data assessment and visualization, supervised and unsupervised machine learning, natural language processing.

FACULTY REQUIREMENTS

COURSE FAMILY

Not Applicable

Course Justification

This course is UC/CSU transferable. It is part of the Python Programming Certificate of Achievement. Data science is a fast-growing field within computer science and as a result, several UCs and other 4-year schools have recently created a bachelor's in data science track to meet the interest and demand in the field. This course helps students prepare for their entrance into a data science major.

Foothill Equivalency

DOES THE COURSE HAVE A FOOTHILL EQUIVALENT?

No

FOOTHILL COURSE ID

Formerly Statement

Course Development Options

BASIC SKILL STATUS (CB08)

Course is not a basic skills course.

GRADE OPTIONS

- Letter Grade
- Pass/No Pass

REPEAT LIMIT

0

Transferability & Gen. Ed. Options

Information below is subject to change. For the official listing of courses, their approval dates, and transfer credit limitations, check the De Anza catalog (by academic year), [ASSIST.ORG \(https://assist.org/\)](https://assist.org/) and [C-ID.NET \(https://c-id.net/\)](https://c-id.net/).

TRANSFERABILITY

Transferable to both UC and CSU

Units and Hours

Summary

MINIMUM CREDIT UNITS 4.5
MAXIMUM CREDIT UNITS 4.5

Weekly Student Hours

Type	In Class	Out of Class
Lecture Hours	4.0	8.0
Laboratory Hours	1.5	0.0

Course Student Hours

COURSE DURATION (WEEKS)

12.0

HOURS PER UNIT DIVISOR

36.0

Course In-Class (Contact) Hours

LECTURE

48.0

LABORATORY

18.0

TOTAL

66.0

Course Out-of-Class Hours

LECTURE

96.0

LABORATORY

0.0

NA

0.0

TOTAL

96.0

Prerequisite(s)

CIS 41A

Corequisite(s)

Advisory(ies)

Limitation(s) on Enrollment

Entrance Skill(s)

General Course Statement(s)

Methods of Instruction

Lecture and visual aids

Discussion of assigned reading Discussion and problem solving performed in class In-class exploration of Internet sites Quiz and examination review performed in class Homework and extended projects Collaborative learning and small group exercises Collaborative projects Laboratory discussion sessions and quizzes that evaluate the proceedings weekly laboratory exercises

Assignments

- A. Reading: required reading from textbook and classnotes
- B. Programs: 6-8 programming homework assignments, several with 100 or more lines of code.

Methods of Evaluation

- A. Evaluation of programming assignments and reports for correctness, use of design principles, documentation and efficiency.
- B. One or more examinations requiring programming ability to develop an algorithm, evaluate code segments, and write code using theories presented in the course.
- C. In-class lab problems, group collaborative problems, exam questions and/or online assignments or tutorials demonstrating the ability to read and analyze code through debugging and/or writing snippets of code.
- D. A final examination requiring programming ability to develop algorithms, evaluate code segments, and write code using theories presented in the course.

Essential Student Materials/Essential College Facilities

Essential Student Materials:

- None.

Essential College Facilities:

- Computer lab with computers running the Python interpreter and Anaconda package

Examples of Primary Texts and References

Author	Title	Publisher	Date/ Edition	ISBN
Igual, Laura and Segui, Santi: Introduction to Data Science, 1st Edition. Springer. ISBN 978-3-319-50017-1. 2017				
Saltz, Jeffrey: An Introduction to Data Science, 1st Edition, Sage Publishing, ISBN: 978-1506377537, 2018				

Examples of Supporting Texts and References

Author	Title	Publisher
Hasti, Trevor: The Elements of Statistical Learning, 2nd Edition, Springer, ISBN: 978-0387848570, 2017		

Learning Outcomes and Objectives

Course Objectives

- Define and describe data science concepts
- Apply mathematics and statistics building blocks
- Apply programming constructs in data science
- Collect data from multiple sources
- Apply data wrangling techniques
- Assess data sets
- Visualize and present data
- Apply a learning framework
- Evaluate supervised learning
- Evaluate unsupervised learning
- Evaluate natural language processing

CSLOs

- Collect, clean, analyze, and visualize data to meet and defend a measured objective.
- Gather data and choose a model to train and tune the machine learning tool and interpret the result.

Outline

- A. Define and describe data science concepts
 - 1. The role of data analytics
 - 2. Machine learning application
- B. Apply mathematics and statistics building blocks
 - 1. Linear algebra - notation, vector and matrix operations
 - 2. Statistics - measurement such as mean, median, mode, standard deviation, outlier, correlation, confidence interval
- C. Apply programming constructs in data science
 - 1. Data Structures
 - a. Array
 - b. Data frame

- 2. Data input / output
- 3. Operators
- 4. Functions
- 5. Plots
- D. Collect data from multiple sources
 - 1. Define objective
 - 2. Web scraping
 - 3. Web API
 - 4. Importing file
 - a. HTML
 - b. text
 - c. CSV
 - d. JSON
- E. Apply data wrangling techniques
 - 1. Joining data
 - 2. Data validation
 - 3. Data cleaning
- F. Assess data sets
 - 1. Exploratory data analysis
 - 2. Data filtering, sorting
 - 3. Searching, retrieving data
 - 4. Time series
- G. Visualize and present data
 - 1. Univariate and multivariate plots
 - 2. Defense of results
- H. Apply a learning framework
 - 1. Loading dataset
 - 2. Learning and predicting
 - 3. Saving models
- I. Evaluate supervised learning
 - 1. Training, predicting
 - 2. Evaluating model success
- J. Evaluate unsupervised learning
 - 1. Clustering
 - 2. Outlier detection
- K. Evaluate natural language processing
 - 1. Rule based and statistical NLP
 - 2. Sentiment analysis

Lab Topics

- A. Present, evaluate, and explain the role and application of data science in modern life.
- B. Solve mathematical and statistical problems with large data sets.
- C. Write code to perform mathematical and statistical work on input data to produce and present the result.
- D. Write code to locate and read data from multiple types of data source.
- E. Write code to join, validate, and clean input data to prepare for analysis.
- F. Write code to filter, search, sort data to arrive at a conclusion on the data trend.
- G. Write code to utilize the appropriate plots to visualize data.
- H. Write code to work with a machine learning model to train and predict data.
- I. Write code to provide input data for a machine learning model for unsupervised learning.
- J. Write code to process text and predict outcome.