

DS 490: DATA SCIENCE CAPSTONE

Degree Programs



Department of Computer Science

COMPUTER SCIENCE BS DEGREE

COMPUTER SCIENCE PROGRAM

- Ranked **#18 overall** by US News & Reports
- 1800+ Computer Science Majors
 - 9 tracks
 - Computational Science and Engineering
 - Computer Graphics and Visualization
 - Database and Information Systems
 - Algorithmic Foundations
 - Machine Intelligence
 - Programming Languages
 - Security
 - Software Engineering
 - Systems Software
- Largest major on Purdue University's campus
- 100% Job placement upon graduation or pursuing graduate degrees

448

DEGREES
GRANTED
LAST
YEAR

211%

UG
INCREASE
OVER 5
YEARS

3.85

YEARS TO
GRADUATE

DATA SCIENCE BS DEGREE

DATA SCIENCE PROGRAM

- Data Science ranks in the top 3 of Glassdoor's Best Jobs in America ranking from 2016-2021
- 402 Data Science Majors (program began 2017)
- 367% increase in growth for Purdue's DS major over 4 years (2018-2021)

“

Data Science is one of the most revolutionary and exciting fields of the 21st century. They always say there are the most opportunities when things just get started. I'm thrilled to be part of this brand new field.

XINYI TAN
Data Science 2020



DATA SCIENCE CERTIFICATE

DATA SCIENCE CERTIFICATE ATTAINABLE TO ANY MAJOR IN AS FEW AS SIX COURSES

- Students gain a foundation in statistical, mathematical and computational techniques for the analysis of large-scale or complex datasets and can apply their skills to different domains.
- Able to interpret results from large-scale data analysis; communicate findings; and recognize and evaluate ethical and social implications of data science.

“

Data science lets me be both creative and technical. It isn't just about plugging data into a program and coming to conclusions based on the output. It's about analyzing it in ways that were never thought of before and using it to solve real-life problems that exist in other industries.

KAVYA NAGALAKUNTA
Computer Science 2020



DATA SCIENCE CONNECTOR MODULES

PREPARES GRADUATE STUDENTS FOR DS RESEARCH IN THEIR DOMAIN

- The College of Science, in a joint effort by the departments of Computer Science, Statistics, and Mathematics, developed an eight (8) credit program.
 - Seven five-week, online, 1-credit STEM modules
 - Covers data science foundational topics.
 - 1-credit Ethics of Data Science module, developed and offered by the Department of Philosophy
- 8 modules are *connector modules* as they teach material from core disciplines that is essential preparation for students to pursue domain-specific, data science research and to enroll in data science graduate courses.



Data Science Course Modules

*INTERESTED IN
APPLYING DATA
SCIENCE TO YOUR
DOMAIN?*

*SEEKING
TO PURSUE
FOUNDATIONAL
DATA SCIENCE
COURSES?*

DATA SCIENCE CONNECTOR MODULES

Data Engineering I
(CS 50023)

Data Engineering II
(CS 50024)

Foundations of CS
(CS 5900FCS)

Foundations of Decision Making
(CS 50025)

Probability and Statistics
(STAT 59800PS)

Numerical Computing for DS
(CS 59000NCDS)

Linear Algebra for DS
(MA 59800)

Ethics for Data Science
(PHIL 29200DL)

***CS COURSES FOR
+ADVANCED
TRAINING***



DEEP LEARNING

ARTIFICIAL
INTELLIGENCE

INFORMATION
RETRIEVAL

DATABASE
SYSTEMS

DATA MINING

STAT MACHINE
LEARNING

NATURAL LANGUAGE
PROCESSING

INTRO TO DATA
VISUALIZATION

DS 490 Data Science Capstone Project (Spring 2023)

Learning outcomes:

After successful completion of this course, a student will be able to:

- Identify the business objectives in a complex real-world problem.
- Develop appropriate data science goals and project plan.
- Implement the appropriate data pre-processing and analysis models.
- Evaluate models and interpret results with respect to the business objectives.
- Present the project outcomes through written reports and digital/oral presentations.

DS 490 Data Science Capstone Project (Spring 2023)

More details:

- We anticipate 100+ students in Spring 2023. Students will work in 4-7 person teams; teamwork is important.
- Teams typically produce excellent products using things they have been taught throughout their four years in DS/CS and often use tools that they have not been taught in class.
- Significant TA support from the CS department.

Do you have a project that involves data analysis?

Let's work together in DS 490 in Spring 2023!

- Students like projects that involve real data that are important for business decisions.
- The project should reach a “culmination point” at the end of the semester.
- Students are OK with preprocessing the data for a while; at some point though they need to work on data analysis. (Iterative refinement is, of course, fine.)
- Some mentoring is useful: meeting once every two to three weeks is beneficial to all parties.
- Short project descriptions are better; somewhat open-ended projects are preferred.
- **Real or “realistic” data is a must.**

CHALLENGES AND OPPORTUNITIES

Data Analytics

- Constant pursuit for “better” clustering/classification/dimensionality reduction approaches.
- *Foundation*: improved algorithms for large-scale optimization.
- Explainable/interpretable algorithms are of paramount importance.
- Algorithmic performance must scale in modern computer architectures (multicore, cloud, GPUs, supercomputers, etc.)
- Fairness has emerged as an important metric for algorithms that have social implications.

CHALLENGES AND OPPORTUNITIES

Privacy

- Sharing raw data is often impossible or very time consuming due to legal constraints.
- *Privacy preserving data mining* is an alternative approach.
- *Share summary statistics of the underlying dataset that allow*
 - *Allow some analyses to be performed*
 - *Not violating the privacy of the underlying dataset.*
- Theory/practice over the past 20 years has explored the tradeoffs between privacy and accuracy of data analytics.

CHALLENGES AND OPPORTUNITIES

Mechanism Design

- The collection of multimodal and complex datasets in, say, the life sciences is done by an intricate ecosystem, involving:
 - commercial enterprises,
 - academic and non-academic researchers,
 - crowdsourcing or other mechanisms to collect individual data.
- Computer Science *has systematically investigated* how to quantify the incentives and rewards within such ecosystems.
- Research on the design of such mechanisms *has not been systematically transferred* to other fields.
- Such advances could potentially result in *improved participation* mechanisms among the general population, as well as underrepresented populations.

THANK YOU!

QUESTIONS?



Department of Computer Science

@PurdueCS



Do you have a project that involves data analysis?

Let's work together in DS 490 in Spring 2023!

- Students like projects that involve real data that are important for business decisions.
- The project should reach a “culmination point” at the end of the semester.
- Students are OK with preprocessing the data for a while; at some point though they need to work on data analysis. (Iterative refinement is, of course, fine.)
- Some mentoring is useful: meeting once every two to three weeks is beneficial to all parties.
- Short project descriptions are better; somewhat open-ended projects are preferred.
- **Real or “realistic” data are a must.**