Installation

Get the required software

1.

Anaconda

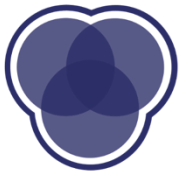
https://www.anaconda.com/distribution/

2.

Open **Terminal (**MacOS**) / Command Prompt (**Windows**),**
Type and enter: jupyter notebook

Lecture 1: Introduction

INFO 1998: Introduction to Machine Learning



CDS Education

Agenda

1. **Meet the Team**: Who are we?
2. **Course Syllabus**: What will we learn?
3. **Introduction**: What is Data Science / Machine Learning?
4. **Course Logistics**: How will we learn?
5. **Getting Started**: Software & Demo

Who are we?

Cornell Data Science

Project Team

- Data Science
- Machine Learning Engineering
- Data Engineering
- Quantitative Finance

Community Outreach

- INFO 1998
- Tech Talks

Course Manager

Who you'll have to bear with

Deniz Bölöni-Turgut

CS '26

db823@cornell.edu

Took INFO 1998 in Spring 2023

Course Staff

Backbone of INFO 1998

- Deniz Bölöni-Turgut
- Mericel Tao
- Srivatsa Kundurthy
- Mahitha Penmetsa
- Cade Jin
- Johann Lee
- Skai Nzeuton
- Minhaj Fahad
- Aarsha Joshi
- Peter Ha
- Daniel Wang
- Samantha Vaca

Getting to know your classmates

Spend 5 minutes going over the following:

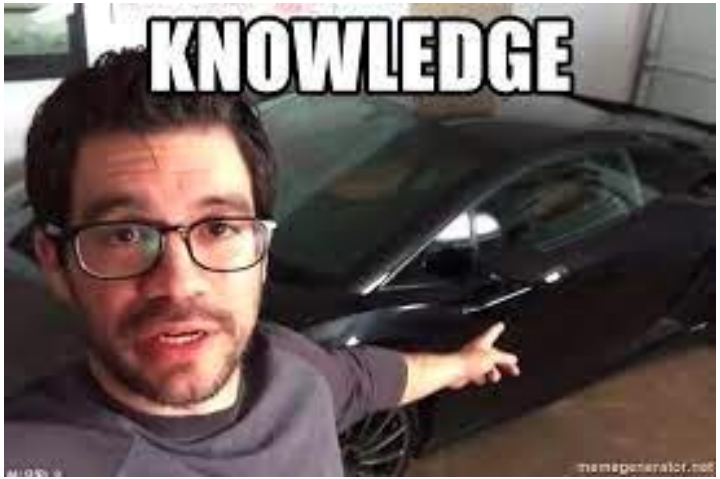
- Name
- Major
- Why you are taking this course
- An application of data science you find interesting

Project Team Website: cornelldata.science
Course Website: cornelldatascience.github.io/info1998/

What Do You Get Out of This?

What you will have accomplished by the end of this?

F·R·I·E·N·D·S



What is Data Science?

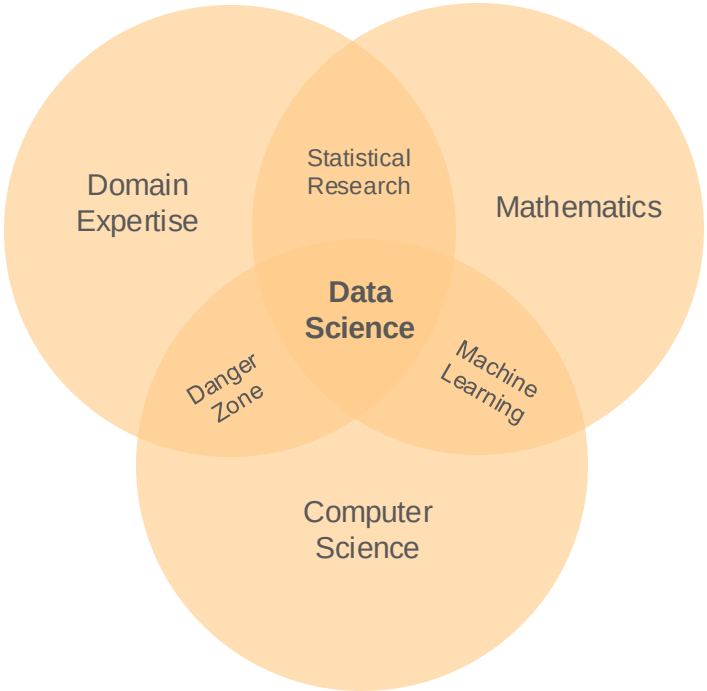
What is Data Science?

“By "Data Science", we mean almost everything that has something to do with data: Collecting, analyzing, modeling..... yet the most important part is its applications --- all sorts of applications.”

Journal of Data Science

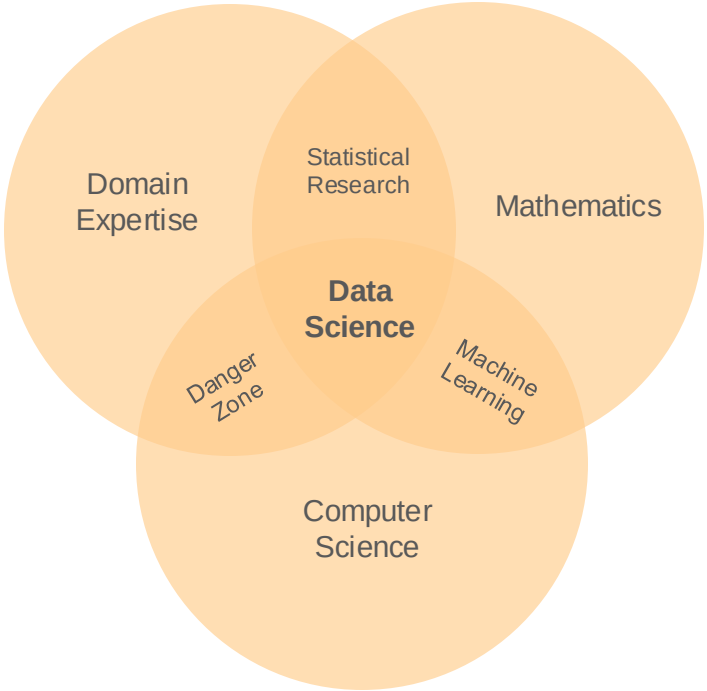
What is Data Science?

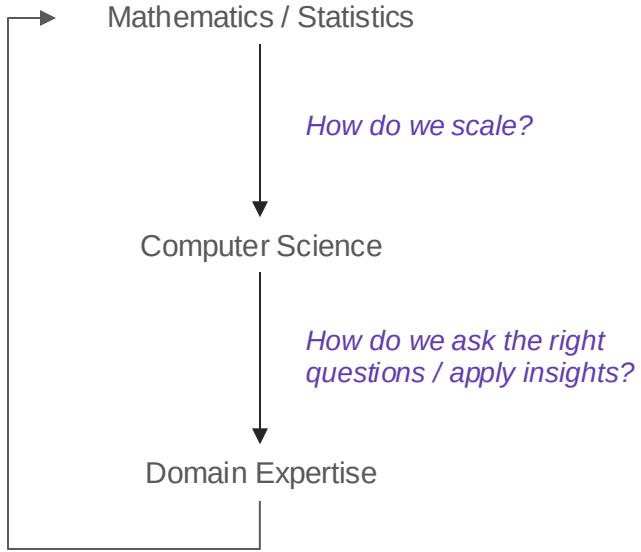
Conway's Data Science Venn Diagram



What is Data Science?

Conway's Data Science Venn Diagram





Data Science ≠ Machine Learning

Applications of Data Science

We'll be back to this slide!

Predictive	Stock Prices
	Netflix Recommendations
Preventive	Medical Diagnosis
	Social Impact Analytics
Real-Time	Digital Advertising
	Autonomous Vehicles

Course Objectives and Syllabus

What you should aim to understand by the end of the course

OBJECTIVES	SYLLABUS
Manipulating Data	Data Manipulation / Visualization *Lectures 1-3*
Communicating Data	
Understanding of ML as a concept	Fundamentals of Machine Learning *Lectures 4-5*
Intuitive understanding of ML models	
Implementation of ML models	Supervised Learning *Lectures 6-8*
Comfort Using Python	
Applications in Industry	Unsupervised Learning *Lecture 9*
Project Experience	

Syllabus is posted on our website.

FAQs

Is this class a good fit for you?

1) Will I become a Data Scientist / Machine Learning Engineer?

No, you will not. The course covers a breadth of concepts, helps build intuitive understanding of some models, but does not dive too deep into the mathematical complexities (since this is a 1000-level course). However, feel free to come to office hours if you're interested in learning more.

2) How much time commitment is this course?

Completely up to you. It's not hard to pass the class if all you want is basic street-fighting machine learning skills, which is fine too – that'll require less than 1 hour per week. If you want to put some more time in and come up with a creative and cool data science project that you can be proud of, that will take more time but is very rewarding!

3) I have no background in CS / Stats – am I underprepared?

Not at all! We'll teach you everything you need to know, but you may have to spend a little more time getting comfortable with Python. A number of non-STEM graduate students have taken this class in the past to understand basics that they could apply to their research. A large number of freshmen also take the course because they're excited to learn more about the field. TL;DR: If you're interested, give it a shot!

Course Logistics

How is the class structured (and graded)?

9 assignments (~1 assignment per lecture) — **55%**

We drop your lowest score!

Project — **40%**

2-3 students

Attendance — **5%**

What you are doing right now ☺

Passing Grade: 70%

Project Details

More details to come on the final project rubric and our website

Pre-Processing and Manipulation

Any necessary cleaning and manipulation of the dataset

Visualizations

At least two different visualizations. Visualizations are clearly visible, clean, well-labeled, and serve a clear purpose for your question(s).

Models

Machine learning models that are chosen wisely, implemented correctly, and give meaningful results. For example, you won't get points if you run a linear regression for a classification problem. If applicable, the results of the models are compared.

Feel free to stay after class or post on Ed to form groups!

Enrollment

Let's get this credit

Fill out by Friday to get a pin.
Also counts for today's attendance!

Enrollment

Let's get this credit

1

Fill out by Friday to get a pin.
https://forms.gle/FNDXZJSkCeAPH5NU7

Enroll in Ed Discussion
https://edstem.org/us/join/QzfhAa
You will be added to CMS over the weekend (after enrollment)

Enrollment

Let's get this credit

1

Fill out by Friday to get a pin.
https://forms.gle/FNDXZJSkCeAPH5NU7

2

Enroll in Ed Discussion
https://edstem.org/us/join/QzfhAa
You will be added to CMS over the weekend

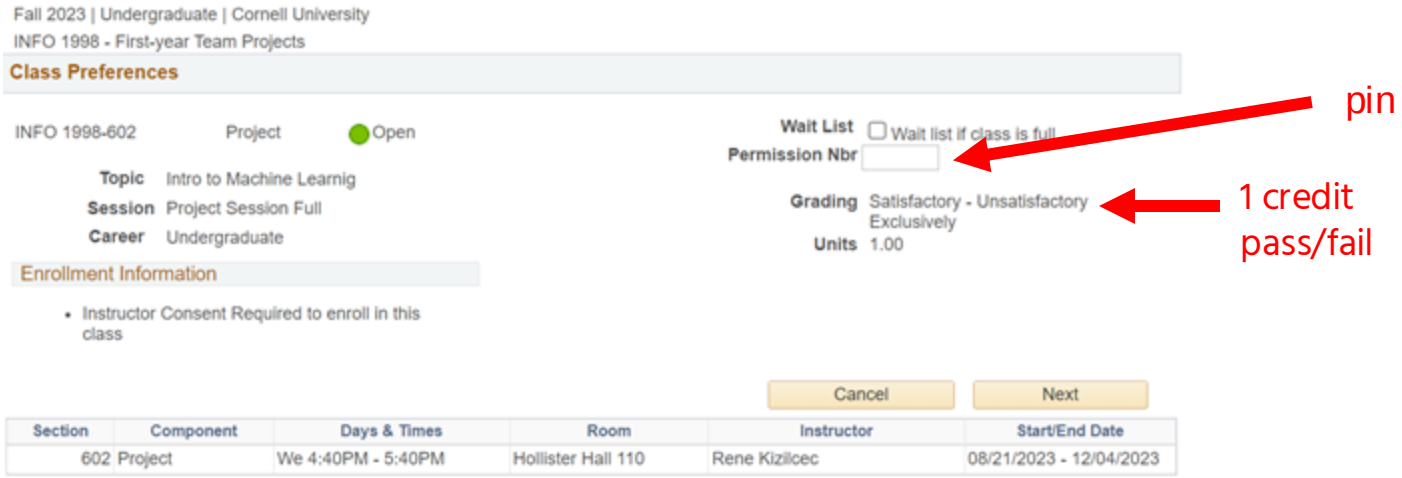
3

Enroll through Student Center after obtaining pins

Enrollment on Student Center

Let's get this credit

- Get enrollment pin via email (sometime next week)
- Add INFO 1998 Section 602 (class # ...) under Rene Kizilcec
- Please try to enroll as soon as possible when you receive your pin



Where can I find course information?

Asking Questions and Course Announcements: Ed Discussion (avoid email if possible)

https://edstem.org/us/join/VmFC2G

Assignment File & Submitting Assignments: CMS

https://cmsx.cs.cornell.edu/

Lecture Recordings, Office Hours Schedule & Assignment Files: Course Website

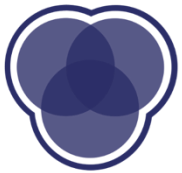
https://cornelldatascience.github.io/info1998/

Jupyter Notebook Demo

Feel free to follow along in Jupyter Notebook
(linked on website)

Next Steps

- **Installation**: Seek help at Office Hours!
- **Assignment 1**: Due at 11:59pm on Friday September 20th, 2024 on **CMSx**.
  .ipynb file is on the website!
- **Enroll on Student Center:** Will receive a pin via Cornell email sometime next week
- **Next Lecture**: Data Manipulation



CDS Education