INFO 1998: Introduction to Machine Learning

CDS Education

We explore, learn, and educate big minds.

Lecture 1: Introduction

INFO 1998: Introduction to Machine Learning

CDS Education

We explore, learn, and educate big minds.

Agenda

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

Who are we?

Cornell Data Science

Project Team

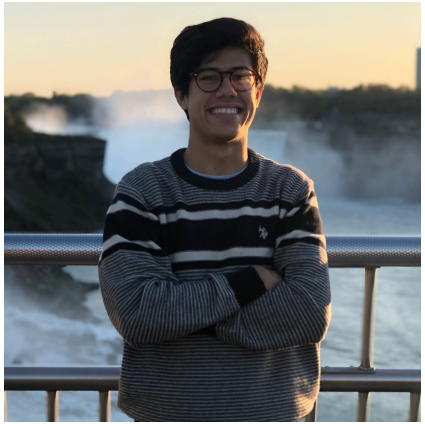
- Intelligent Systems
- Insights
- Data Engineering
- Algorithmic Trading

Community Outreach

- Education
  - INFO 1998
  - Workshops
  - Online Tutorials

Course Manager

Who you'll have to bear with



Camilo Cedeno-Tobon

ORIE '21

cc2459@cornell.edu

Took INFO 1998 in Fall 2018

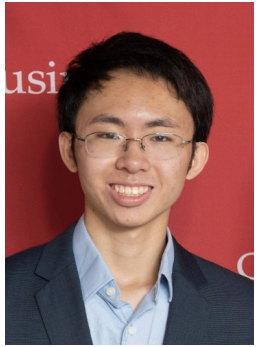
Course Instructors

Backbone of INFO 1998

Sam Cobado
ECE ‘22

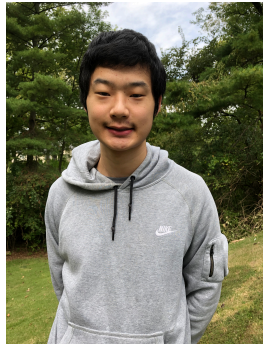
Dylan Tsai
CS mEng

Emily Weed
Stats ‘22



Jerry Sun
CS ‘23

Raye Liu
ORIE ‘22



Kevin Jiang
ORIE ‘23

What is Data Science?

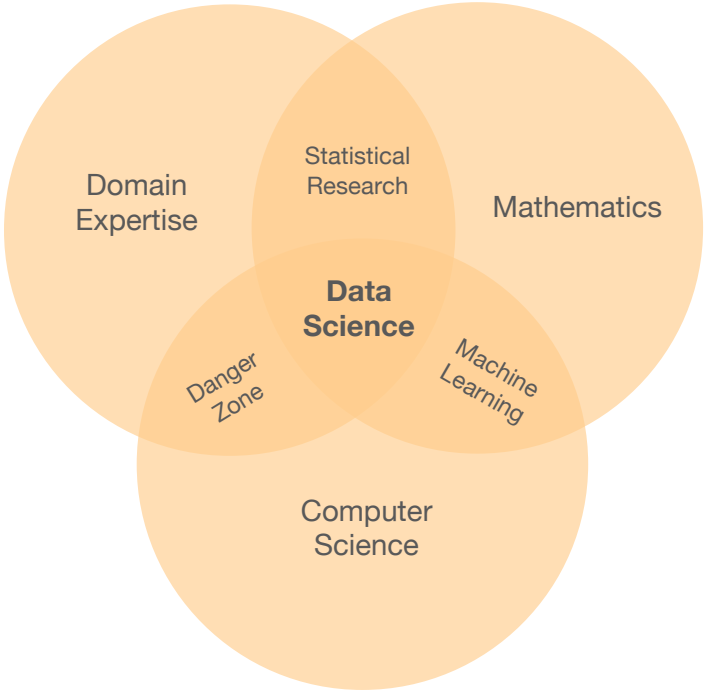
What is Data Science?

“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

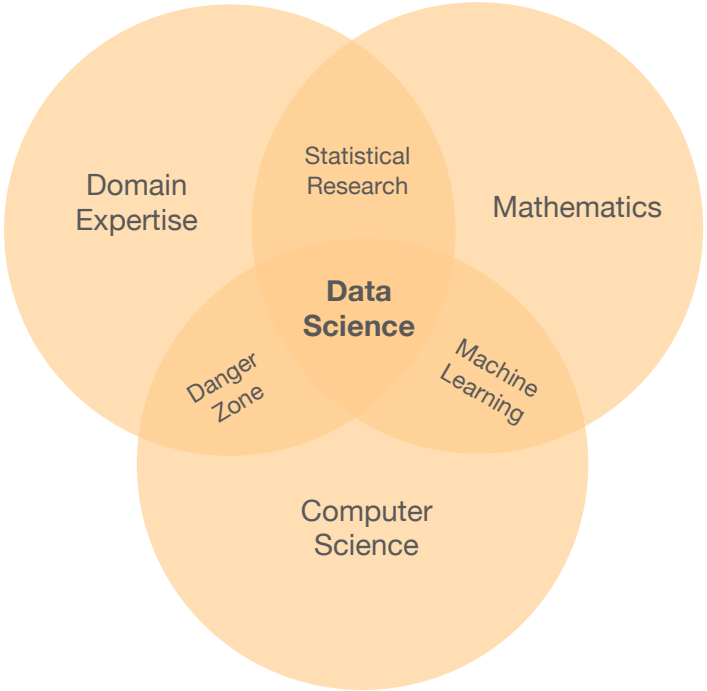
What is Data Science?

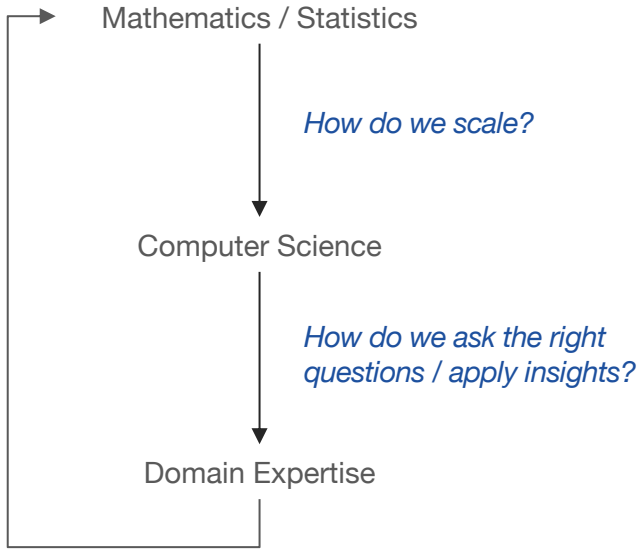
Conway's Data Science Venn Diagram



What is Data Science?

Conway's Data Science Venn Diagram





Data Science ≠ Machine Learning

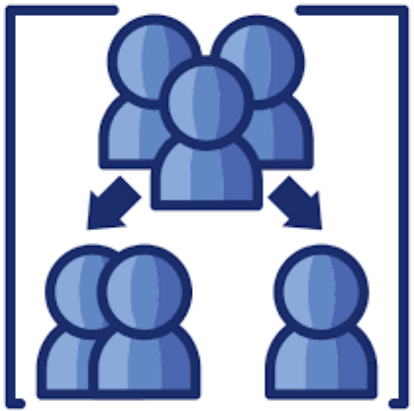
Applications of Data Science

We'll be back to this slide!

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

Getting to know your classmates

Breakout Rooms



Spend 5 minutes going over the following:

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

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 Piazza under ”resources”

Sample Final Projects

“0 – 100, Real Quick” - Drake

(1) Determining indicators for a candidate’s success in Canadian Elections

Kevin Zhou, Jerry Sun

(2) Predicting the amount of solar radiation the Earth gets

Raye Liu, Justin Lee

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?

Depends. If you want to have a strong command over the material so that you can get a head start in this field, you will have to read a little more and be prepared to spend time with our TAs to go over concepts in more depth. If you want to acquire just street-fighting machine learning skills, that's fine too – it'll require little more than 1 hour per week.

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!

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_fall20

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_fall20

2

Enroll in Piazza
(INFO 1998)
You will be added to CMS
over the weekend

Enrollment

Let's get this credit

1

Fill out
tiny.cc/info1998_fall20

2

Enroll in Piazza
(INFO 1998)
You will be added to CMS over the weekend

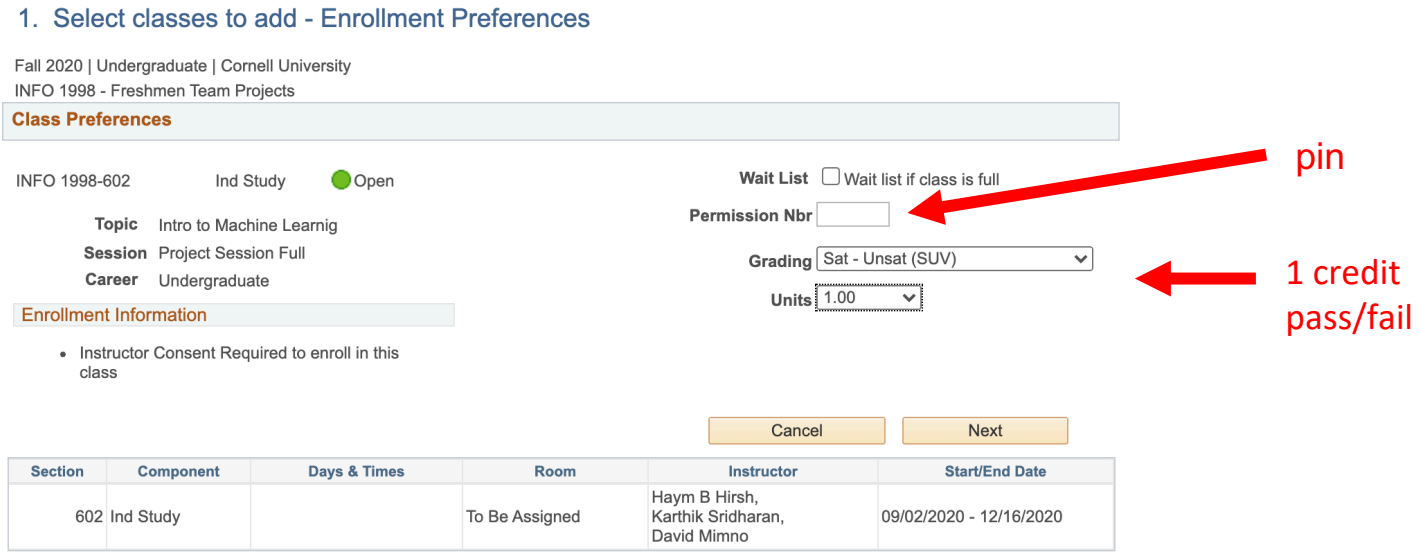
3

Enroll through Student Center after obtaining pins

Enrollment on Student Center

Let's get this credit

- **Get enrollment pin via email (by end of Thursday 09/24)**
- **Add INFO 1998 Section 602 under David Mimno**
- **Deadline to enroll is 10/02**



Course Logistics

How is the class structured (and graded)?

10 assignments (1 assignment per lecture) *Drop lowest score!*	**50%**
Mid-semester Group Project *2-3 students*	**15%**
Final Group Project *2-3 students*	**35%**

Passing Grade: 70%

Where can I find course information?

Asking Questions / Assignments / Lecture Slides: Piazza

https://www.piazza.com/cornell/fall2020/info1998

Submitting Assignments: CMS

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

Lecture Recordings: Course Website

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

Getting Started

Get the required software

Jupyter Notebook

- Documentation (Code + Visuals)
- Supports Python, Julia, etc.
- Easy to share

Python (3)

- Easy to learn, Readable
- Industry Standard
- Great documentation, online resources

Installation

Get the required software

1.

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

2.

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

Demo

Next Steps

- **Installation**: Seek help at Office Hours!
- **Assignment 1**: Due at 5:30pm on Wednesday, Sep 30, 2020 on CMS (Will be enrolled soon!)
- **Python Workshop:** Please fill out interest form (http://tiny.cc/python_workshop)
- **Enroll on Student Center:** Wil receive a pin through your Cornell email later tonight
- **Next Lecture**: Data Manipulation

CDS Education

We explore, learn, and educate big minds.