

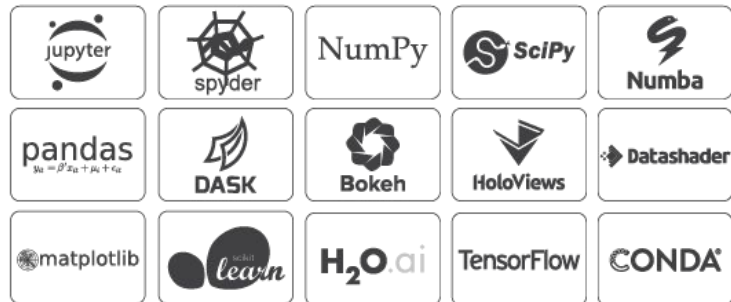
Installation

Get the required software

1.



Anaconda



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

2.

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

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. **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

Online Tutorials

Course Manager

Who you'll have to bear with



Varun Gande

ORIE '24

vg262@cornell.edu

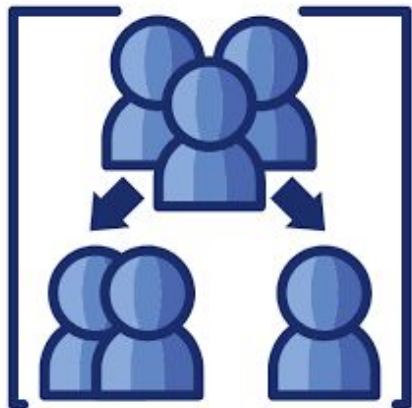
Took INFO 1998 in Spring 2020

Course Staff

Backbone of INFO 1998

- Vivian Chen
- Vincent Fong
- Eric Guo
- Kelly Lu
- Neha Kulshreshtha
- Jacob Mayourian
- Max Wang
- Mahitha Penmetsa
- Everett Lee
- Koji Kimura

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

cornelldata.science
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

Things I do when I have to learn.



- Learn
- Think about learning
- Find a million excuses why I don't have the time to learn

SCIENCE or PEOPLE



What is Data Science?



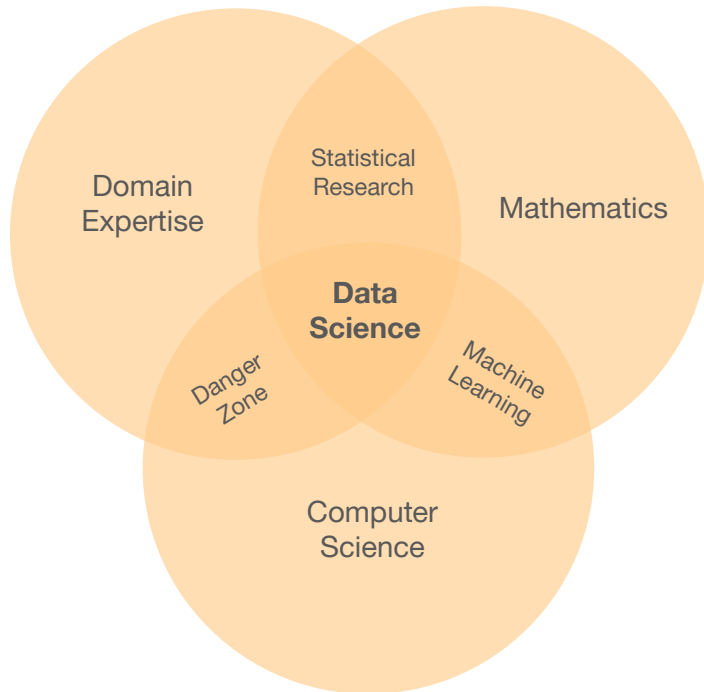
What is Data Science?

“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

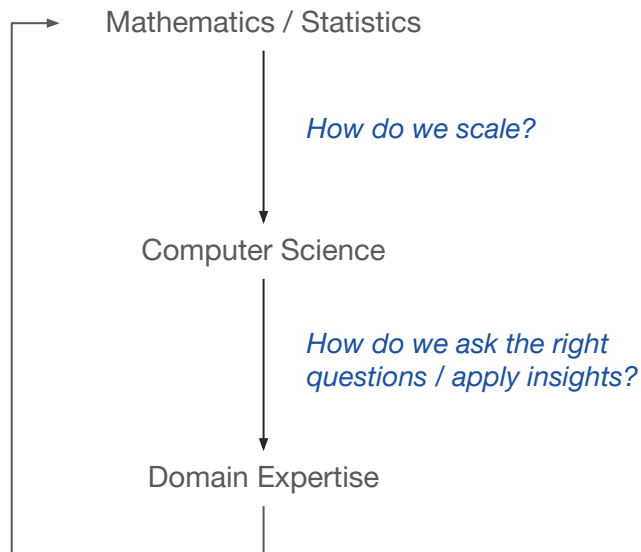
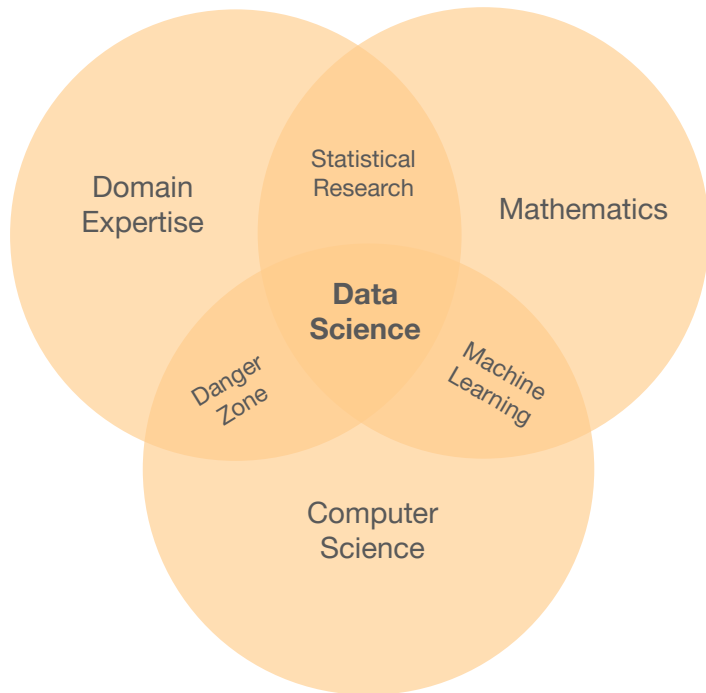
What is Data Science?

Conway's Data Science Venn Diagram



What is Data Science?

Conway's Data Science Venn Diagram



Data Science \neq 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 <i>Lectures 1-3</i>
Communicating Data	
Understanding of ML as a concept	Fundamentals of Machine Learning <i>Lectures 4-5</i>
Intuitive understanding of ML models	
Implementation of ML models	Supervised Learning <i>Lectures 6-8</i>
Comfort Using Python	
Applications in Industry	Unsupervised Learning <i>Lecture 9</i>
Project Experience	

Syllabus is posted on our website under “syllabus”



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)

We drop your lowest score!

60%

Project

2-3 students

40%

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 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 to form groups!



Sample Final Projects UPDATE

"0 – 100, Real Quick" - Drake

(1) Predicting Heart Failure

Fernando Celaya, Ming DeMers, Marcus Posey

(2) Predicting the Price of Used Cars

Grant Rineheimer, Benjamin Tang, Dylan Tom

Enrollment

Let's get this credit



Fill out by Friday to get a pin
<https://forms.gle/QtXdDQhKhyQWewuUA>

Enrollment

Let's get this credit

1

Fill out by Friday to get a pin

<https://forms.gle/QtXdDQhKhyQWewuU>

A



Enroll in Ed Discussion

<https://edstem.org/us/join/znT86d>

You will be added to CMS over the weekend

Enrollment

Let's get this credit

1

Fill out by Friday to get a pin

<https://forms.gle/QtXdDQhKhyQWewuU>

A

2

Enroll in Ed Discussion

<https://edstem.org/us/join/znT86d>

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 # 13536) under Rene Kizilcec**
- **Please try to enroll as soon as possible when you receive your pin**

1. Select classes to add - Enrollment Preferences

Spring 2023 | Undergraduate | Cornell University

INFO 1998 - Freshmen Team Projects

Class Preferences

INFO 1998-602 Project ☒ Open

Topic Intro to Machine Learnig

Session Project Session Full

Career Undergraduate

Wait List ☐ Wait list if class

Permission Nbr

Grading Sat - Unsat (SUV)

Units 1.00

Enrollment Information

- Instructor Consent Required to enroll in this class

Cancel

Previous

Next

Section	Component	Days & Times	Room	Instructor	Start/End Date
602 Project		We 4:30PM - 5:30PM	Hollister Hall B14	Rene Kizilcec	01/23/2023 - 05/09/2023

NOTES

Class Notes Staff will be handling enrollment during the add/drop period in January. Please fill out this form: https://docs.google.com/forms/d/e/1FAIpQLSVcSX8VWQ-oEuaEDSVLK0w_gnL0EqSLAcJBDD7paahnbFR6A/viewform

pin

1 credit
pass/fail

Where can I find course information?

Asking Questions: Ed Discussion (avoid email if possible)

<https://edstem.org/us/join/znT86d>

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



Next Steps

- **Installation:** Seek help at Office Hours!
- **Assignment 1:** Due at 11:59pm on Wednesday, Feb 22, 2023 on CMS. File template is on the website!
- **Enroll on Student Center:** Will receive a pin via Cornell email sometime next week
- **Next Lecture:** Data Manipulation



CDS Education

We explore, learn, and educate big minds.