

Statistics, Machine Learning, and Data Mining

Syllabus for Ma322 Section 802 Fall 2026

Info

3 Credits

Mon 3:00pm - 6:00pm

Room: B825

Instructor Information

calvin_williamson@fitnyc.edu

office: B831 Science and Math

office hours: M 1-3, W 11-12, R 12-1, or appointment

Description

This is an introduction to statistical techniques for machine learning and data mining. It emphasizes mathematical methods and computer applications related to automated learning for prediction, classification, knowledge discovery and forecasting in modern data science. Special emphasis will be given to the collection, mining, and analysis of massive data sets. (G2: Mathematics) Prerequisite(s): MA 222 and mathematic proficiency (see beginning of Mathematics section)

Outcomes

Upon completion of this course students will be able to:

1. Describe the concepts of machine learning and identify examples of its use in data science.
2. Employ statistical software to collect data, create training and test sets, and perform predictions.
3. Create regression models for predicting outcome variables in terms of predictors.
4. Explain the contributions of Google in understanding web scale data and the structure of the internet.
5. Identify the characteristics of massive data sets and describe the tools needed to analyze them.
6. Analyze decision tree models and display them with appropriate graphics.
7. Use recommendation systems software and understand how it makes suggestions based on similarity measures.

8. Perform classifications for data sets using nearest neighbor and probabilistic algorithms.
9. Collect text data and use text mining software to perform sentiment analysis.

Course Materials

We will be using Google Colab and Google Sheets for all work in this course. Since these are web-based applications there is NO OTHER SOFTWARE required for the course besides a web browser.

Topics

Regression

- Simple Regression
- Multiple Regression
- Applications
- Conjoint Analysis

Introduction to Python

- Google Colab Notebook
- Using LLM as Coding Assistant
- Calculations
- Variables
- DataTypes
- Lists
- Dictionaries
- Functions
- Dataframes
- Strings

Machine Learning using Agents

- Classification, Accuracy
- Training, Testing
- Decision Trees

Agents

- Github
- Codespaces
- Terminals
- OpenCode, Claude Code, Codex

- Agentic Software Development

Agents and Databases

- Relational Databases
- SQL
- Dolt Databases
- Agentic Data Science

Evaluation

Your grade will come from these parts:

- Quizzes (88%)
- In Class/Homework Credits (12%)

Each of these parts is described in more detail below

Quizzes

Your quiz grade will come from 5 quizzes, roughly covering 2 or 3 weeks of material each. The quizzes are 30 minutes each and are usually 5 or 6 questions each. These quizzes are with no notes, no internet, no phone, no software, no AI tools. Pen and paper and calculator only. They are some multiple choice, some short answer, some true false.

In Class/Homework Credits (1-3 per class)

These are credits you obtain for demonstrating you have completed assigned problems.

Some of these will come from homework assignments that you show me at the beginning of the class, some of these will come from in class assignments that are done during class and you show as you complete them. You will earn a single credit for each successful assignment completion. You must be in attendance to earn these problem credits. I decide if the work is complete enough to receive credit for.

The In Class/Homework can only be turned in by showing it to me. There is not a way to turn these assignments by emailing them to me or submitting in Brightspace, or any other digital way of submitting them.

You show them to me and I mark it as complete when you are in class.

You may miss up to 2 classes and you can still show me the assignments from those classes when you are in class again. But you may not do this more than 2 weeks worth.

Some homeworks will be counted for a grade, but not always.

Midterm, Final Exam

There is NO MIDTERM and NO FINAL EXAM in this course.

AI Policy

Your use of AI in this course is not restricted. Generally all assignments use AI in some form, and you are free to use AI tools in whatever way you find helpful on them.

For some assignments though the AI is meant to be used in a particular way, with a specific tool, prompt, or technique, and working that way is part of what the assignment is teaching. When an assignment calls for that, please try to adhere to it.

The quizzes are the exception. Those are pen and paper with no notes, no internet, no phone, no software, and no AI tools, so use the AI to help you learn the material rather than in place of learning it.