Machine Learning for the Business Analyst

Effective: 1/18/2021 (SUBJECT TO CHANGE; visit course site regularly for updates)

01/18/2021-05/06/2021
Section 01, CRN 27780, T/R 11:00am-12:15pm, Jepson G28
Section 02, CRN 27781, T/R 12:45pm-2:00pm, Jepson G24

Zoom Link:
https://urichmond.zoom.us/my/profmattson?pwd=K0NGKzJSZzdYUXdGYk9PM0JvQ296QT09

Academic Calendar: https://registrar.richmond.edu/planning/index.html

Instructor: Thomas Mattson
Office: Queally 381, Office Phone: (804) 287-6895
Office Hours: Tuesday & Thursday 3:00pm-5:00pm and by appointment
Note: I am generally available to meet during normal business hours (except when I am teaching). I will also make sure that I am available on Zoom during times that are convenient for my remote students who are multiple time zones away from Richmond.
E-mail: tmattson@richmond.edu

Course Overview

Machine learning is the process of using a variety of different algorithms (i.e., processes or sets of rules) to teach a computer how to make accurate predictions about some phenomenon based on the patterns and anomalies in the data that get fed into the different algorithms. As new data are fed into the machine learning algorithms, the machine learning models morph and can make more (or possibly less) accurate predictions based on the updated data. A machine learning model is the output of a "training" data set that is used to teach or fine-tune the model's predictive power. Machine learning models may be online (continuously being updated) or offline (updated periodically). We see machine learning models at work in our everyday lives on social media (targeted advertisements), in our email inboxes (detecting spam from non-spam emails), and on electronic commerce sites (product recommendations). Organizations use machine learning models to predict (among many others) market movements, employee turnover, accounting fraud, changing customer preferences, and price erosion.

Figure 1. Machine Learning Process

Using machine learning requires a different way to approach solving or uncovering business opportunities and threats. Figure 1 displays the machine learning process, which starts with data collection and does not conform to the standard scientific method. With machine learning, we solve problems in a bottom-up manner starting with the data instead of a top-down approach

starting with management theories, a-priori variables, experience, and expertise. This approach requires letting the machine learning model do the heavy analytical lifting to solve the business problem, which managers then combine with their contextual knowledge to make (hopefully) better (or at least more informed) decisions. This problem-solving approach represents a shift for people familiar with thinking through functional steps and the traditional scientific method. It often takes a change in organizational culture or philosophy to trust that the machine learning model will produce accurate answers to their business problems. There are certainly both pros (speed and less prone to cognitive biases) and cons (not everything that counts can be counted and "training" data may not be representative of realty) to using this approach to aid in managerial decision making.

Machine learning is a component of the broader concept of artificial intelligence (i.e., computer systems that can perform tasks that usually require human intelligence). Figure 2 displays a graphic depicting where machine learning fits within artificial intelligence.
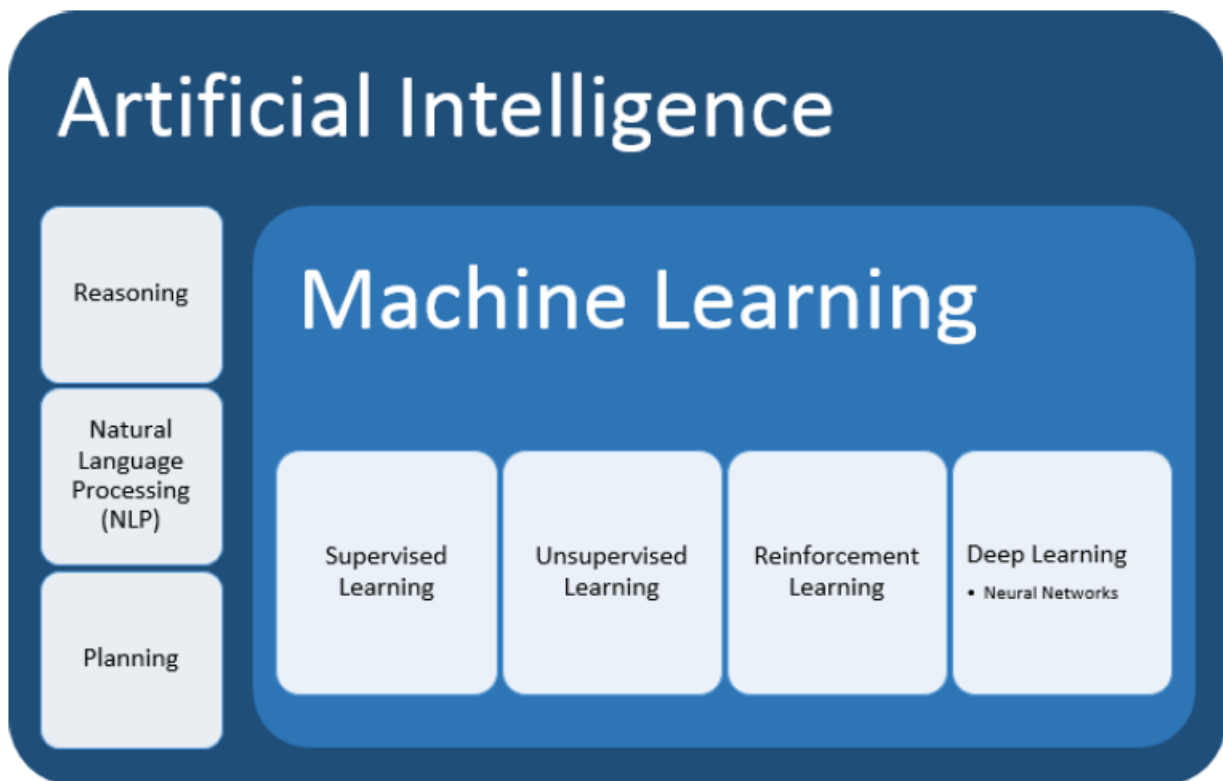

Figure 2. Artificial Intelligence and Machine Learning

In this class, we will focus primarily on supervised and unsupervised machine learning models. Time permitting, we will discuss and implement different algorithms for reinforcement and deep learning. Figure 3 displays the differences between supervised and unsupervised machine learning. In the case of supervised learning, we train a machine learning model to make predictions using known inputs (features) and known outputs (targets). Once the model has seen enough examples in its "training" data set, it can predict a probable output based on a series of known inputs with previously unseen data. With unsupervised learning, we train a machine learning model to find patterns in the "training" data using a set of inputs (features) without providing the algorithm with known or desired outcomes (targets).
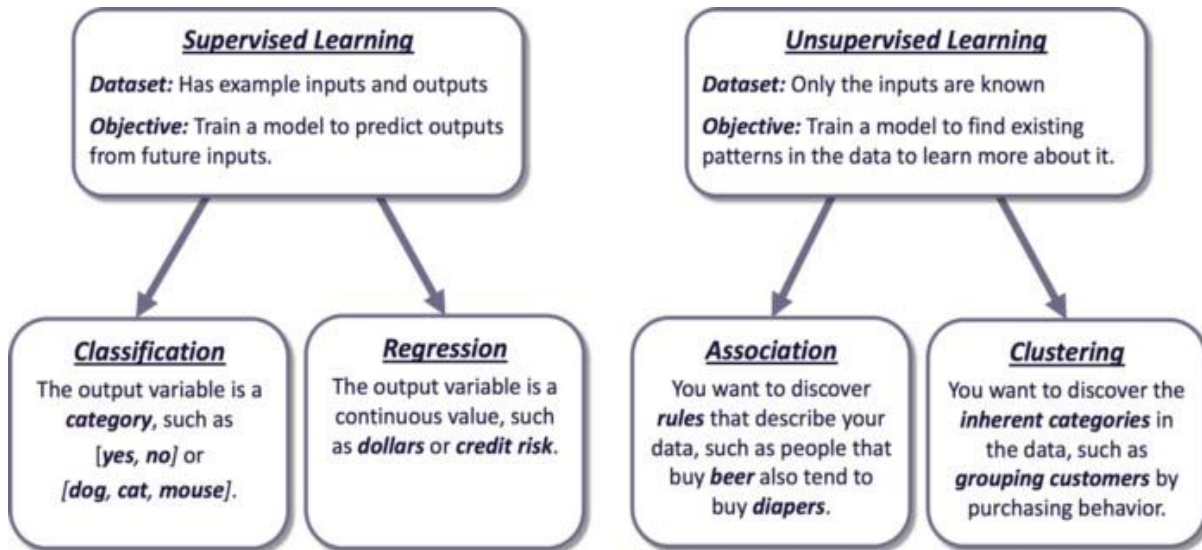
Figure 3. Supervised and Unsupervised Machine Learning

There are many different applications and programming languages that are used to implement machine learning models. In this class, we will use Python to develop our machine learning models because it is currently widely used across different organizations and industries. Additionally, there are a variety of free machine learning frameworks and libraries available for anyone to use. However, the processes and concepts that you will learn in this class should be applicable to other languages and environments if your future company does not use Python.

Learning Objectives

In this class, students will analyze and think critically about how to solve complex business problems using machine learning. In doing so, students will produce relevant machine learning solutions that have practical relevance for business managers and other organizational stakeholders. Additionally, students will effectively communicate their machine learning solutions to non-technical users or decision makers. Finally, students will think critically about the ethics associated with developing and deploying these types of machine learning models.

More specifically, students will learn:

- Python fundamentals. This course is a not a Python course per se. You will learn enough Python such that you can successfully perform machine learning using Python.
- To use Python to implement both supervised and unsupervised machine learning algorithms.
- To understand conceptually the pros and cons of using machine learning models in organizations.
- To further develop your critical and analytical thinking skills as it applies to data analytics.
- Further develop your sense of integrity and learn how to be an honorable business or data analyst when applying machine learning.

Software

All of the video tutorials will be developed on Windows machines. However, each piece of software should work on a MAC but the workflow and dialog boxes might differ on a MAC relative to a Windows machine so the instructional materials might not be 100% identical if you are following along on a MAC. Therefore, my suggestion is to install bootcamp or parallels on your MAC if you want to use your personal machine (see https://www.parallels.com/landingpage/pd/windows-on-mac/ to help you get started setting your MAC to run Windows). The University help desk should also be able to assist you in setting up your MAC to run Windows.

We will use Python3 in this course (https://www.python.org/downloads/). Please note that Python3 is different from Python2 so please make sure to download and install Python3. You can install Anaconda if you want but I will be using the base Python3 installation in all of my instructional materials. The Anaconda installation will result in a slightly different set-up in PyCharm.

We will use the PyCharm IDE (integrated development environment) to write our Python scripts in this course (https://www.jetbrains.com/pycharm/download/#section=windows). You do not need to purchase the professional edition of the software because the free community version has all of the necessary features you will need to complete the course assignments. I have provided the link to the Windows version of this IDE because all of the instructional materials will be developed on the Windows platform. However, there is a MAC version of the PyCharm IDE if you want to use your native MAC environment.

We will also use MongoDB as the backend for several projects and mini-projects in this course. You do not have to install MongoDB on your local machine because we have a server that we will all use. However, if you want to install MongoDB on your local laptop, you can install the community edition (https://www.mongodb.com/community). The free version is sufficient for everything that we will do in this course.

General Course Policies & Guidelines

Class preparation & debriefing: To be successful in this course you should expect to devote 10-14 hours each week in class, reading and studying the material, and preparing assignments. Learning how to program requires significant repetition to learn. How much repetition is required will vary from individual to individual. As such, preparation will require some combination of repeating the in-class exercises, completing the homework problems, and reading the appropriate references. The in-class mini-cases/examples will require you to have the homework completed before the start of the class period in order to successfully complete the in-class activities.

The instructor as resource: "To teach" means to "pass on knowledge," while "to learn" means "to acquire knowledge and skills." As course instructor, my job is to teach you what I know and understand about python and machine learning by organizing the course around important topics, key instructional materials and assignments while acting as a facilitator, resource, and guide. You can ask me questions and ask me for extra help and I will do my best to assist you, but it is your job to learn (even when learning is hard work)! Much of the learning will require you to repeat the instructional tutorials multiple times.

I took this position because I enjoy teaching. I genuinely care about you and your progress in the class. If you have a problem, complaint, comment, concern, etc., **please** schedule an appointment or drop in during open office hours. If something is not working for you, don't wait until the very end of the semester to speak up. Come in and speak with me so we can work together to maximize your experience in this course.

Expectations for student behavior: I expect each student to demonstrate **respect** for his or her fellow students individually and as a class, to me as your instructor, to the Robins School of Business as a place of learning – and to himself or herself. In addition to the typical ways we should be respectful (e.g., language use, politeness, cooperation, openness to new ideas, etc.), being respectful means **no cheating and no plagiarism**. To be clear, **cheating** includes, but is not limited to, copying someone else's work, with or without their knowledge, and turning it in as your own work. **Plagiarism** includes, but is not limited to, copying intellectual property from others (on the web, in books, etc.) and presenting it as your own work, without proper citation. If I identify instances of cheating or plagiarism, ***the students involved will receive a failing grade (F) for the class and instances of cheating or plagiarism will be reported to the Honor Council for possible further sanctions***. Make sure you familiarize yourself with Richmond's honor code (http://studentdevelopment.richmond.edu/student-handbook/honor/the-honor-code.html).   Not   knowing   the policies and procedures related to my course and the University of Richmond is not an excuse to violate those policies and procedures. *When in doubt, don't do it and ask me for clarification*.

Class Participation & Attendance: My assumption is that all students who are taking my class are adults. If you, as an adult, choose to miss class on a given day, I assume that 1) you understand that you are missing a component of the class, 2) you have decided that whatever you have chosen to attend in lieu of class is more important than what is happening in class that day, and 3) you are willing to accept any consequences for missing what happens in class that day (lecture material, interaction with your classmates, classroom presentations, exams, and so on).

Religious Observance Policy: Students needing to miss class because of religious observance should contact me within the first two weeks of the semester to discuss the absence. The University's full religious observance policy may be found here (http://registrar.richmond.edu/services/policies/religiousobsv.html).

Other Resources Available to Students:

If you experience difficulties in this course, do not hesitate to consult with me. There are also other resources that can support you in your efforts to meet course requirements.

Academic Skills Center (asc.richmond.edu): Assists students in assessing their academic strengths and weaknesses; honing their academic skills through teaching effective test preparation, critical reading and thinking, information conceptualization, concentration, and related techniques; working on specific subject areas (e.g., calculus, chemistry, accounting, etc.); and encouraging campus and community involvement. Tutors will be available virtually. The on-call peer-tutors available for these appointments are listed in the Box file: On-Call Online Tutors (https://richmond.box.com/s/dpe37chr2zodr3o1amtj8omjk72v2ktb). **Email** Roger Mancastroppa (rmancast@richmond.edu) **and** Hope Walton (hwalton@richmond.edu) **for appointments in academic and life skills to request a Zoom conference.**

Boatwright Library Research Librarians: (library.richmond.edu/help/ask/ or 289-8876): Research librarians help students with all steps of their research, from identifying or narrowing a topic, to locating, accessing, evaluating, and citing information resources. Librarians support students in their classes across the curriculum and provide library instruction, tutorials, research guides, and individual help. All research support will be provided online or by appointment and students can contact a librarian for help via email (library@richmond.edu), text (804-277-9ASK), chat, or Zoom (by appointment).

Career Services: (careerservices.richmond.edu or 289-8547): Can assist you in exploring your interests and abilities, choosing a major or course of study, connecting with internships and jobs, and investigating graduate and professional school options. We encourage you to schedule an appointment with a career advisor early in your time at UR.

Counseling and Psychological Services (caps.richmond.edu or 289-8119): Assists currently enrolled, full-time, degree-seeking students in improving their mental health and well-being, and in handling challenges that may impede their growth and development. Services include brief consultations, short-term counseling and psychotherapy, skills-building classes, crisis intervention, psychiatric consultation, and related services.

Disability Services (disability.richmond.edu) The Office of Disability Services works to ensure that qualified students with a disability (whether incoming or current) are provided with reasonable accommodations that enable students to participate fully in activities, programs, services and benefits provided to all students. Please let your professors know as soon as possible if you have an accommodation that requires academic coordination and planning.

Speech Center (speech.richmond.edu or 289-6409): Assists with preparation and practice in the pursuit of   excellence in public expression. Recording, playback, coaching and critique sessions offered by teams of student consultants trained to assist in developing ideas, arranging key points for more effective organization, improving style and delivery, and handling multimedia aids for individual and group presentations. Remote practice sessions can be arranged; we look forward to meeting your public speaking needs.

Writing Center (writing.richmond.edu or 289-8263): Assists writers at all levels of experience, across all majors. Students can schedule appointments with trained writing consultants who offer friendly critiques of written work.

Assignments & Deliverables

Homework Assignments & In-class Exercises: Learning how to program in Python and how to apply machine learning algorithms in Python require continuous practice. Each week, you will be given a series of programming tasks (and other conceptual assignments). Some of those tasks will be completed in-class and other tasks will be completed outside of class. You may certainly discuss these activities with other students but each student is responsible for doing the work themselves (after some optional discussion with your fellow classmates). To be clear, this type of discussion does ***NOT*** mean that one student does the work and then emails the Python script to the other student. Instead, it means that you may sit side-by-side and help each other solve the specific problems. Each student must write their own code (or their own written answers) but they may get help from another student. Emailing files to one another for submission is an honor code violation in this class. I will only collect these assignments if I suspect that the class as a whole is not completing them. If I collect them and you did not complete them, you will get negative points assigned to your grade.

Tests/Assessments: There will be three assessments that will be completely ***individually***. The format of each assessment will vary depending on the topics being assessed.

Project: There will be a ***group*** project where you will apply the knowledge and skills you learned in this class to a business problem.

Adding up the points

	Points	%
Test #1 – Python Core Constructs	25	25%
Test #2 – Specialized Python Packages/Modules	25	25%
Test #3 – Machine Learning	25	25%
Group Project	25	25%
Total	100	

NOTE: Failure to complete the homework assignments (when I collect them) will result in anywhere between -1 and -5 points towards your final grade.

A >=93%	A- >=90% & <93%	B+ >=87% & <90%
B >=83% & <87%	B- >=80% & <83%	C+ >=77% & <80%
C >=73% & <77%	C- >=70% & <73%	D+ >=67% & <70%
D >=63% & <67%	D- >=60% & <63%	F < 60%

TENTATIVE Course Schedule

Part 1 – Course Introduction & Introduction to Python	
Weeks 1 to 5	***Topics***: 1. General Introduction to Machine Learning 2. Set up the Python environment in PyCharm, write your first program, and execute your first program (in the command prompt and in PyCharm). 3. Core Python Constructs - variables, conditions, statements, iteration, functions, lists, tuples, dictionaries, and general i/o 4. Version control with Github
Test for Part 1 tentatively scheduled at the end of Week 5	
Part 2 – Specialized Python Modules & Packages	
Weeks 6 to 9	***Topics***: 1. Pandas, SciPy, Numpy, & Statsmodels 2. MongoDB & PyMongo 3. Reading & writing to/from different databases 4. Visualizing data using Matplotlib, Seaborn, and Bokeh (time permitting) 5. Other specialized packages (time permitting)
Test for Part 2 tentatively scheduled at the end of Week 9	
Part 3 – Supervised & Unsupervised Machine Learning	
Weeks 10 to 12	***Topics***: 1. Machine Learning Terminology and Concepts 2. K-Nearest Neighbors 3. Gradient Descent 4. Logistic Regression 5. Regression Decision Trees 6. Ensemble Models (time permitting) 7. Support Vector Machines (time permitting)
Weeks 13 to 14	***Topics***: 1. K-means 2. Principal Component Analysis
Test for Part 3 tentatively scheduled at the start of Week 15	
Week 15	Project Work
Project due on the date of the final exam	