Woods College of Advancing Studies, Boston College

ADAN8888 — Applied Analytics Project

Instructor Information

Instructor: Nurtekin Savas **Email Address:** savasnu@bc.edu

Course Description

Our objective, in this course, is to simulate the life of a model developer in the industry enabling students to complete a machine learning project from end to end. We will cover identifying a problem that can be solved using machine learning, articulating the business intent, building the business case, finding the data from a public repository that will solve the problem, data analysis to explore the dataset, feature engineering that is appropriate for the problem area and model, developing models, analysis of models, evaluation of models, analysis and consideration of ethical / fairness / bias issues, and developing a model monitoring strategy.

Course Format

This course will be held asynchronously via Canvas.

Textbooks (with ISBN), Readings and Learning Resources (Required):

All articles/handouts will be available in Canvas unless otherwise noted.

Required Text

None

Other equipment / material requirements (optional)

None

Textbooks & Readings (Recommended)

None

Recommended Text (if applicable)

None

Multimedia Resources

I will add videos describing the weekly objectives and discussing the why of each module

Course Outcomes

By the end of this course, you will be able to:

1. Define a modeling / machine learning problem, articulate the intent and build a business case
2. Find the right dataset that will help solve the problem, analyze the dataset, clean the data, engineer features appropriate for the problem area
3. Build multiple types of models and optimize hyper-parameters of the models. Evaluate each model and pick the winning model
4. Package model for deployment and build a monitoring plan
5. Analyze ethical, fairness and bias issues present in the data and models. Build a model monitoring plan

6. Evaluate the quality of an end to end modeling work and its artifacts via peer review

Assessments & Grade Weighting

You will need to develop and update three artifacts / outputs each week to demonstrate your progress.

1. A Jupyter notebook containing all code, which runs error-free

2. A Github repo

3. A written report documenting what you've done and accomplished for the week

4. A written report providing feedback to one student's artifacts

Weight	Assessments	Meets Outcome #	Due
24%	Jupyter notebook that contains the code to produce outputs in an error free manner	All learning outcomes	Weeks 2-12
8%	Creating a GitHub project repo and merging artifacts to the GitHub project repo	All learning outcomes	Weeks 1-12
27%	Written report documenting the analysis, outputs and findings	All learning outcomes	Weeks 1-12
8%	Code output that shows executed code with relevant outputs	All learning outcomes	Weeks 2-12

17%	Final Jupyter notebook, GitHub project repo and written report	All learning outcomes	Week 13
6%	Written feedback report on a peer's work	Learning outcome #5	Week 14
10%	Discussion	All learning outcomes	Weeks 1-14

Below you will find more details about each of these activities or assignments.

Jupyter Notebook

You are expected to develop a Jupyter notebook to complete each week's learning objectives. This notebook should contain only relevant code for the week's activities. The code should be structured in a logical manner with appropriate notes embedded into the notebook. It is expected that the notebook will run error free and produce all outputs relevant for the week.

Code Output

You are expected to run the code each week and provide the output of the code.

Github Repo

You are expected to create and maintain a well-organized project repo in Github. Each week the final version of the code in the Jupyter notebook should be merged into your Github repo.

Report

You are expected to write a report that documents what you did, why you did it and how you did it with only relevant insights, findings and analysis to meet the objectives for the week.

Discussion

You are expected to participate in weekly discussions and respond to questions for each.

Final docs

In week 13 you are expected to combine all week's code, code output, docs and merge these final artifacts to GitHub repos.

Peer feedback

In week 14 you are expected to review a peer's artifacts (code, code output, GitHub repo, and doc from week 13) and provide written feedback.

Rubrics

You will find all activity or assignment rubrics in Canvas.

Grading The course will be graded out of 100 points. **A** = 94-100; **A-** = 90-93; **B+** = 87-89; **B** = 84-86; **B-** = 80-83; **C+** = 77-79; **C** = 73-76; **C-** = 70-72; **D+** = 67-69; **D** = 63-66; **D-** = 60-62; **F** = 59 and below.

Course Assignments

It is expected that you will spend 9-12 hours per week on assignments. Specifics of weekly assignments are listed below. Please note that some weeks will require more time and some weeks less.

Course Outline

Week	Weekly Outcomes By the end of this week, you will be	Learning Resources	Assessments

	able to:		
Week 1 Identify the problem statement and dataset	1.Identify an opportunity that can be solved with machine learning 2.Articulate the value of solving the problem and develop a business case 3.Initiate project in GitHub and set up programming environment in JupyterHub	You can find a comprehensive list of general sample resources each week in Canvas.	· Written report · Create GitHub account · Set up Jupyter Hub environment
Week 2 Ingest and explore the dataset	1. Ingest data into an IDE 2. Identify what you're predicting (i.e., dependent variable) 3. Identify what you're using as predictors (I.e., independent variables) 4. Explore the variables in the dataset	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These resources show you how data is ingested using python. Feel free to learn from these sources but you may need something different for your particular dataset.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 3 Perform exploratory data analysis	1.Split your dataset into training, validation and test datasets 2.Perform EDA on your datasets 3.Identify problems, opportunities and pre-processing needed	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show you how to split data into training, validation and test datasets and conduct EDA using python. Feel free to learn from these sources but you may need something different for your particular dataset.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 4	1.Make data model ready by	You can find a comprehensive list of general	· Written report · Jupyter

Make data model ready	preprocessing your data based on findings from Week 3 and 4 2.Perform preprocessing across training, validation and test datasets	sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show you how to preprocess datasets using python. Feel free to learn from these sources but you may need something different for your particular dataset.	notebook · Code output · Merge all artifacts into GitHub
Week 5 Engineer features	1.Create and engineer new features 2.Augment your dataset with new data 3.Reduce dimensions of your dataset 4.Create final training, validation and test datasets that will be used in modeling, evaluation and testing	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show you how to split data into training, validation and test datasets using python. Feel free to learn from these sources but you may need something different for your particular dataset.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 6 Develop 1st modeling approach (simple, the baseline)	1.Build a very simple set of models appropriate for your problem and dataset 2.Tune hyper-parameters of the model 3.Select model evaluation metrics, evaluate variations and pick the winning model	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show you how to train a model, do hyper parameter search/tuning, and model evaluation using python. Feel free to learn from these sources but you may need something different for your particular choice of model and/or dataset.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 7	1.Build a more complex set of	You can find a comprehensive list of general	· Written report · Jupyter

Develop 2nd modeling approach (more complex)	models appropriate for your problem and dataset 2.Tune hyper-parameters of the model 3.Select model evaluation metrics, evaluate variations and pick the winning model	sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources discuss model complexity, overfitting and the bias-variance tradeoff in models. Feel free to learn from these sources but you may need something different for your particular choice of model and/or dataset.	notebook · Code output · Merge all artifacts into GitHub
Week 8 Develop 3rd modeling approach (even more complex)	1. Build another set of models appropriate for your problem and dataset 2. Tune hyper-parameters of the model 3. Select model evaluation metrics, evaluate variations and pick the winning model	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources discuss model complexity, overfitting and the bias-variance tradeoff in models. Feel free to learn from these sources but you may need something different for your particular choice of model and/or dataset.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 9 Select the winning model	1.Evaluate multiple models 2.Select the best model 3.Calculate performance on the test dataset	None	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 10 Data Centric AI	1.Improve your model by improving the data	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub

Week 11 Explain the model, analyze risk, bias and ethical consideratio ns	1.Explain your model by understanding feature importance and prediction outcomes 2.Identify model risks 3. Identify and quantify bias in your input dataset and model output 4.Identify and measure bias	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show you how to extract important features from certain types of models using python, discuss model explainability, AI ethics, bias in models, and model risk. Feel free to learn from these sources but you may need something different for your particular choice of model, dataset or problem statement.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub Identify a partner peer for your feedback report in Weeks 14-15 (not-graded but required) · Auto matching peers without a partner (non-graded activity)
Week 12 Save and package your model for deployment. Build your model monitoring plan	1.Save and deploy your model 2.Build a monitoring plan	You can find a comprehensive list of general sample resources each week in Canvas. You are encouraged to find others on the internet. These sample resources show how to save, load and serialize a model using python as well as provide an overview of data drift, concept drift and model monitoring. Feel free to learn from these sources but you may need something different for your particular choice of model.	· Written report · Jupyter notebook · Code output · Merge all artifacts into GitHub
Week 13 Bring it all together	1.To put all weeks together and develop an end to end modeling project with all the artifacts	None	· Final written report · Final Jupyter notebook · Final code output · Final merge all artifacts into GitHub

Week 14 Review a peer's work and provide feedback	1. Review a peer's end to end modeling work and related artifacts 2.Provide written feedback to a peer's work and artifacts	None	· Go through the other person's artifacts – no deliverable · Written peer feedback report

Classroom Policies

Participation

Participating in class is an important component of learning. Students are expected to participate and complete all assignments, and assessments. While this is an asynchronous course, participating in online class discussion boards is an important component of learning. These discussion boards will take the place of our in-class interactions and should be used as both a space to provide opinions and thoughts, while also posing questions or asking for clarification on theories and concepts.

Consistent with BC's commitment to creating a learning environment that is respectful of persons of differing backgrounds, we believe that every reasonable effort should be made to allow members of the university community to observe their religious holidays without jeopardizing their academic status. Students are responsible for reviewing course syllabi as soon as possible, and for communicating with the instructor promptly regarding any possible conflicts with observed religious holidays. Students are responsible for completing all class requirements for days missed due to conflicts with religious holidays.

Submitting Papers

All writing assignments, code, code outputs and GitHub submissions are to be submitted through Canvas no later than

11:59 PM on Sunday. Please submit your outputs in the following formats:

- Written report: .pdf format
- Jupyter code: .ipynb format
- Jupyter code output: .ipynb format or .pdf format
- GitHub: provide repo url in that week's discussion board

Late Work Policies Assignments or discussions submitted after 11:59 PM on the due date will result in a penalty of 25% per day until the assignment is submitted. For example, if you submit your assignment 1 day late your maximum grade will be 75%, 2 days late will result in a maximum grade of 50%, 3 days late will result in a maximum grade of 25%. After three days, the assignment or discussion will no longer be accepted. In the event of illness, emergency, other extenuating circumstances, or school-sanctioned events, please contact me as soon as possible, providing written verification of the excused absence. Makeup/late assignment submissions must be completed no later than 5 days after the original assignment due date. There are no exceptions to this policy.

Communication You may contact me by email and I will respond within 24 hours.

Boston College Policies

Academic Integrity I have no tolerance for cheating in any form. Such will earn the student an automatic zero. Academic misconduct includes (but is not limited to): using material (including code or code blocks) from the internet without citing it; plagiarizing any part of work done by someone else; and submitting substantially similar work for two courses without consent. It is your responsibility to familiarize yourself with Boston College policy on academic integrity. If you have any questions, always consult with me.

Accommodations Boston College is committed to providing accommodations to students, faculty, staff, and visitors with disabilities. Specific documentation from the appropriate office is required for students seeking accommodation in Woods College courses. Advanced notice and formal registration with the appropriate office are required to facilitate this process. There are two separate offices at BC that coordinate services for students with disabilities:

- The Connors Family Learning Center (CFLC) coordinates services for students with LD and ADHD.
- The Disabilities Services Office (DSO) coordinates services for all other disabilities.

Find out more about BC's commitment to accessibility at www.bc.edu/sites/accessibility.