Northeastern University
College of Engineering

INFO 7375
Special Topics in
Artificial Intelligence Engineering
Computational Skepticism

Course Information
Special Topics in Artificial Intelligence Engineering Computational Skepticism
INFO 7375
Summer 2020
4 Credit Hours
Location (*100% online through BB Collab)*

Instructor Information
Professor: Nik Bear Brown
Email: ni.brown@neu.edu
Office: 505A Dana Hall

Course Time
3:20 pm - 5:00 pm TR

100% online through BB Collab

Technical/Course Materials Requirements
Readings to be given weekly in class.
Need laptop or access to a computer.

The course GitHub (for all lectures, assignments and projects):

https://github.com/nikbearbrown/INFO_7375

nikbearbrown YouTube channel

Over the course of the semester I'll be making and putting additional data science and machine learning related video's on my YouTube channel.

https://www.youtube.com/user/nikbearbrown

Slack

Join the Slack - https://join.slack.com/t/neuaiskunkworks/shared_invite/enQtNzQyNDg1MjgzNjM0LTYxMWRhMWViMWIxMzUxMTg0YjI4YTQ2NTQyOWM1MmNkOThkYWI0MWU4Y2MyZjA2Njg2Y2Y0YjRjNjQwNWY3MDk

Then join channel #INFO_7375

Course Description

Trust but verify,' is a proverb that should be a mantra for the age of artificial intelligence. Despite their widespread adoption, machine learning models remain mostly black boxes. In spite of this, many who use machine learning to make critical predictions in domains such as finance, telecommunication, healthcare, and many other domains don't fully understand how machine learning models make their predictions.

In this research, seminar we do research in Computational Skepticism, that is, building systems to answer the question "Why Should I Trust AI?". Computational Skepticism a process of obtaining knowledge through automating systematic doubt and continual testing. The word Skepticism comes from the Greek skeptomai, or to search for alternative possibilities.

As engineering research, the focus is building systems that can be used to transform untrustworthy models into a trustworthy ones. To accomplish this there are a number of component subprojects using techniques from probability, deep learning, reinforcement learning, machine learning, and data visualization. Students are expected to pick a single subproject within the first two weeks. All students should have a strong programming background. The sub-projects typically require some knowledge of either probability, deep learning, reinforcement learning, machine learning or data visualization. The approval of a student's choice of subproject will depend on the student's interest and background. The current list of projects and associated code is kept at the INFO 7375 GitHub https://github.com/nikbearbrown/INFO_7375

Students present their research every two to three weeks.

What is Computational Skepticism?

Evidence, as broadly construed, is anything presented in support or opposition of an assertion. Creating an automated skepticism pipeline is intended to do three things: 1) accumulate as much evidence for an assertion or question as it can in an automated way, 2) store and search that data and 3) present and visualize that data in an understandable manner.

There are many forms of evidence in machine learning, all of which are active areas of research. The goal of Computational Skepticism is to automate the best practices in all of these areas of research into a single, intuitive framework. The forms of automated evidence gathering to be studied include: data quality and completeness, bias and fairness , AutoML, model interpretability, causal inference, counterfactual models, deep learning pipelines (AutoDL), time-series pipelines (AutoTS), feature engineering pipelines AutoFE), autovisualization (AutoViz), reinforcement learning pipelines (AutoRL), evidence knowledge graphs (EKG) and many more.

Course Prerequisite
Approval of the instructor.

Student Learning/Course Outcomes (SLOs)
This course is a research seminar that is intended to guide students through the conceptualization, planning, and execution of a major original project in Artificial Intelligence, Machine Learning, Data Visualization or Reinforcement Learning.

The course outcomes of the research seminar are:

1. To provide students with additional training in research design, research methods, and effective writing;
2. To guide students through the process of conducting original research and analysis, leading to the production of substantial, professional-quality papers. This course also helps students recognize common research mistakes and biases, while learning more about what constitutes strong research; and
3. To help students develop their abilities to constructively critique and contribute to the work of others.

Students will be collaboratively writing the first book ever on Computational Skepticism. Small groups of students will collaborate on writing a chapter. Two students have already started on their chapter on model interpretability, so you can see what the beginnings of this process looks like here https://maheshwarappa-a.gitbook.io/ads/

Once completed the Computational Skepticism book will be available for free online and published with an ISBN through the Banataba project through a publishing site such as https://www.Blurb.com.

Attendance Policy
Participation in discussions is an important aspect on the class. It is important that both students and instructional staff help foster an environment in which students feel safe asking questions, posing their opinions, and sharing their work for critique. If at any time you feel this environment is being threatened—by other students, the TA, or the professor—speak up and make your concerns heard. If you feel uncomfortable broaching this topic with the professor, you should feel free to voice your concerns to the Dean's office. Students are expected to complete course readings, participate in class discussions or other learning activities during the unit, and complete written assignments for each unit during the time of that unit.

It is understood that there might be one week when active participation in ongoing class conversations and learning activities might be delayed. Beyond one week time, if there is an absence or lateness in participation (1) faculty must be notified in advance; (2) grades will be adjusted accordingly.

Collaboration Policy
Students are strongly encouraged to collaborate through discussing strategies for completing assignments, talking about the readings before class, and studying for the exams. However, all work that you turn in to me with your name on it must be in your own words or coded in your own style. Directly copied code or text from any other source MUST be cited. In any case, you must write up

your solutions, in your own words. Furthermore, if you did collaborate on any problem, you must clearly list all of the collaborators in your submission. Handing in the same work for more than one course without explicit permission is forbidden.

Feel free to discuss general strategies, but any written work or code should be your own, in your own words/style. If you have collaborated on ideas leading up to the final solution, give each other credit on what you turn in, clearly labeling who contributed what ideas. Individuals should be able to explain the function of every aspect of group-produced work. Not understanding what plagiarism is does not constitute an excuse for committing it. You should familiarize yourself with the University's policies on academic dishonesty at the beginning of the semester. If you have any doubts whatsoever about whether you are breaking the rules – ask!

Any submitted work violating the collaboration policies WILL BE GIVEN A ZERO even if "by mistake." Multiple mistakes will be sent to OSCCR for disciplinary review.

To reiterate: plagiarism and cheating are strictly forbidden. No excuses, no exceptions. All incidents of plagiarism and cheating will be sent to OSCCR for disciplinary review.

Late Work Policy

Students must submit assignments by the deadline in the time zone noted on BlackBoard. Students must communicate with the faculty prior to the deadline if they anticipate work will be submitted late.

Work submitted late without prior communication with faculty will be deducted 10% for each day late.

Grading/Evaluation Standards

Students are evaluated based on their performance on assignments, performance on exams, and both the execution and presentation of a final project. If a particular grade is required in this class to satisfy any external criteria—including, but not limited to, employment opportunities, visa maintenance, scholarships, and financial aid—it is the student's responsibility to earn that grade by working consistently throughout the semester. Grades will not be changed based on student need, nor will extra credit opportunities be provided to an individual student without being made available to the entire class.

Grade Scale

The following breakdown will be used for determining the final course grade:

Assignment	Percent of Total Grade
Paricipation	50%
Mid-term Project	20%
Final Project	30%

* Note that the assignments, presentations and drafts related to the research project go to that score rather than the programming assignments. I expect to use the following grading scale at the end of the semester. You should not expect a curve to be applied; but I reserve the right to use one.

Score	Grade
93 – 100	A
90 – 92	A-
88 – 89	B+
83 – 87	B
80 – 82	B-
78 – 79	C+
73 – 77	C
70 – 72	C-
60 – 69	D
<60	F

Scores in-between grades. For example, 82.5 or 92.3 will be decided based on the exams.

* Note the score is calculated using the grading rubric and IS NOT the average of the assignments that is displayed by BlackBoard.

Course Schedule

This is a seminar class. The literature is read and presented every week. Students present their research every two to three weeks. The beginning of each week will introduce new theory. The focus for Summer 2020 Parts 1 through IV below. Advanced students have the option of working on any of the parts below.

Companies that have applications and research related to Computational Skepticism will be invited from time to time.

Part 0 Assertions and Questions

A preface to the course discusses how to formulate questions and assertions.

Part I Data

The first part discusses understanding data quality, bias, and predictive value so that automated pipelines can be built that assess the quality of a dataset and its appropriateness to answer an assertion. Feature engineering is also discussed. Technqiues include descriptive statsitics, data auditing, exploratory data analysis (EDA), resampling methods.

Part II Bias and Fairness

Discrimination is the unequal treatment of individuals or groups. While algorithms are blind, unintentional unfairness often creeps into the decisions they suggest, particularly in "black box" models. This part discusses how bias can be introduced into the machine learning pipeline, what it means for a decision to be fair, and methods to remove bias and ensure fairness.

Part III Models

The third part discusses building automated pipelines that build models to answer an assertion, evaluating the best models for a given purpose and selecting the most appropriate parsimonious models. The focus is on Automated machine learning (AutoML) is the process of automating the process of feature selection, algorithm selection, hyperparameter optimization, metric selection and creating stacked ensembles.

Technqiues include feature selection, algorithm selection, hyperparameter optimization, metric selection and stacked ensembles.

The base algorithms included are Distributed Random Forest (DRF), Extremely Randomized Trees (XRT), Generalized Linear Model (GLM), Generalized Additive Model (GAM), Gradient Boosting Machine (GBM), XGBoost, and Simple Deep Learning (MLP Neural Networks)

Part IV Model Interpretability

The fourth part discusses building automated pipelines that allow a human to understand the logic and process that a model uses to answer an assertion. This is model interpretability which refers to how easy it is for humans to understand the processes a model uses to arrive at its outcomes.

Technqiues include individual conditional expectation (ICE), leave-one-covariance (LOCO), local feature importance, partial dependency plots, tree-based feature importance, standardized coefficient importance, accumulated local effects (ALE) plots and Shapley values. .

The output of a model interpretability pipeline is as follows:

A Model Schematic Diagram that visualizes all of he steps that each model uses to arrive at its outcomes.

Plots that show the outcome of model interpretability algorithms such as feature importance. partial dependency plots, etc.

A Feature Knowledge Graph with compares illustrates feature importance and interrelationships.

A Data Sensitivity Graph which exposes the effect of adding noise the data on the robustness of a model and the sensitity of individual features.

Advanced Topics

Advanced Topics (These topics will not be presented as part of a structured lecture but advanced students are encouraged to present the literature in the student presentations related to their research topic). I will be looking for experts in these areas of research. The presentations by experts won't necessarily follow any particular order but be scheduled according to the external researchers schedules.

Part V Causal Inference

The fifth part causal inference, in the conext of builing automated pipeline for understanding causation. Causal inference is the process of drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect. The main difference between causal inference and inference of association is that the former analyzes the response of the effect variable when the cause is changed. In this part we discuss causal methods as compared to traditional statistical methods.

Part VI Counterfactual Models

The sixth part discusses building automated pipelines that build counterfactual simulations and use causal methods to ask "what if" questions. Technqiues include causal inference, agent-based modeling and reinforcement learning.

Part VII Deep Learning Pipeline (AutoDL)

The seventh part is an extension of AutoML discussed in part two to use more soptisticated deep learning models like CNNs, RNNs, Gauge-equivariant convolutional neural networks (Gauge CNNs) or other deep learning models.

Part VIII Time-Series Pipeline (AutoTS)

The eigth part is an extension of AutoML discussed in part two to use time series models like ARIMA, VARMA, RNNs, fbProphet, etc.

Part IX Feature Engineering Pipeline (AutoFE)

The ninth part is an extension of AutoML for automated feature extraction.

Part X Autovisualization (AutoVIZ)

The tenth part is the integration of machine learning and data visualization to automatically rank and generate relevant plots for a data set.

Part XI Reinforcement Learning Pipeline (AutoRL)

The eleventh part discusses building automated pipelines that answer optimization questions. The model behind reinforcement learning presupposes one has some form of signal, such as the state of a game board, or the sensors attached to a self-driving car. The central problem that reinforcement learning is intended to solve what is the optimal action to take, given a signal and some objective or reward. For example, what move to make to do well in a game, or what actions to take to not crash a car and get to some destination within any rules of the road.

The AutoRL project is to build automated pipelines for reinforcement learning.

Part XII Evidence Knowledge Graphs (EKG)

The twelth part discusses building Knowledge Graphs. A Knowledge Graph is a network database using information gathered from a variety of sources to present information on a topic or thing and its relations with other things.

For example, a knowledge graph could not only collect facts about an algorithm such as its hyperparameters as one would get in documentation but the distribution of hyperparameter values extracted from its data sources.

It can so "similar" algorithms and which algorithms are used for particular problems. Google used its Knowledge Graph for answering "roughly one-third" of the 100 billion monthly searches Google processed in May 2016.

Knowledge Graphs allow for the addition of "wisdom of the crowds" based evidence into a skeptical framework. Look at the references of the GitHub page under the topics Knowledge Graphs and Collective Intelligence for more information.

Academic Integrity

A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to http://www.northeastern.edu/osccr/academic-integrity-policy/ to access the full academic integrity policy.

Student Accommodations

Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit http://www.northeastern.edu/drc/getting-started-with-the-drc/.

Library Services

The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for Education specific resources, visit http://subjectguides.lib.neu.edu/edresearch.

Diversity and Inclusion

Northeastern University is committed to equal opportunity, affirmative action, diversity and social justice while building a climate of inclusion on and beyond campus. In the classroom, member of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration and an awareness of global perspectives on social justice.

Please visit http://www.northeastern.edu/oidi/ for complete information on Diversity and Inclusion

TITLE IX

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty and staff.

In case of an emergency, please call 911.

Please visit www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.