ADEC7430.01 Big Data Econometrics, Course Credits: 3
Woods College of Advancing Studies, Boston College
Summer 2022, Session 1, May 18–July 1, 2022

Instructor Information:

Name: Haydar Evren
BC E-mail: evrenh@bc.edu
Best way to contact the instructor is via e-mail
Office Hours: By appointment - virtual via Zoom
Class Location: Online

Course Description:

This course demonstrates how to merge economic data analysis and applied econometric tools with the most common machine learning techniques, as the rapid advancement of computational methods provides unprecedented opportunities for understanding "big data". This course will provide a hands-on experience with the terminology, technology and methodologies behind machine learning with economic applications in marketing, finance, healthcare and other areas. The main topics covered in this course include: advanced regression techniques, resampling methods, model selection and regularization, classification models (logistic regression, Naïve Bayes, discriminant analysis, k-nearest neighbors, neural networks), tree-based methods, support vector machines, and unsupervised learning (principal components analysis and clustering). Students will apply both supervised and unsupervised machine learning techniques to solve various economics-related problems with real-world data sets.

Course Delivery:

The format of this course is online and asynchronous.

Textbook (Required):

- An Introduction to Statistical Learning, with Applications in R, by G. James, D. Witten, T. Hastie, and R. Tibshirani. (Digital version available for free download at https://www.statlearning.com/)

Textbooks & Readings (Recommended):

- Applied Predictive Modeling, by M. Kuhn and K. Johnson
- The Elements of Statistical Learning, by T. Hastie, R. Tibshirani, and J. Friedman
- Learning from Data: A Short Course, by Y. Abu-Mostafa, M. Magdon-Ismail, and H. Lin
- Machine Learning: A Probabilistic Perspective, by K. Murphy
- Pattern Recognition and Machine Learning, C. Bishop
- Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, by J. Kelleher, B. Mac Namee, and A. D'Arcy
- Introduction to Machine Learning, 3rd Edition, by E. Alpaydin
- Data Mining and Business Analytics with R, by J. Ledolter

Required Software

The primary software environment is the R statistical programming language, which can be downloaded for free from http://www.r-project. RStudio is the recommended interface for the R statistical programming language software, which can also be downloaded for free at http://www.rstudio.org.

Canvas

Canvas is the Learning Management System (LMS) at Boston College, designed to help faculty and students share ideas, collaborate on assignments, discuss course readings and materials, submit assignments, and much more - all online. Your course will make significant use of Canvas this semester; you should familiarize yourself with this important tool. For more information and training resources for using Canvas, click here.

In the case of any technical difficulties or concerns, please contact canvas@bc.edu or 617-552-HELP (4357) for immediate assistance. Canvas requires particular computer specifications and wifi access. It is important that you plan accordingly.

Course Outcomes

By the end of this course:

1. You will demonstrate a practical understanding of the key theoretical concepts of modern computational/analytic methods across cultural settings and will learn the impact of culture, gender, and age in machine learning as demonstrated by the analysis of real-world data sets.
2. You will demonstrate ethical data usage and data understanding pertaining to the selection of appropriate types and combinations of big data econometric models given particular business situations as demonstrated by the identification of appropriate machine learning methods to find relationships and structure in data with and without specific output variable(s).
3. You will gain an advanced level, practical knowledge of machine learning methods to build predictive models and discover patterns in data for more informed business decision-making, as demonstrated by the successful completion of quizzes, programming assignments, and projects, and contributions to class discussions.
4. You will be able to effectively use a statistical programming software package, as demonstrated by the implementation of various big data econometric techniques using R to develop analytic solutions by building and deploying specific machine learning models based on real-world, practical business problems to learn how to enhance business capabilities and extend the value of existing data, transforming data into knowledge.

Course Approach

The course is conducted entirely online (fully asynchronous) via Canvas. Each week, the student will complete assigned reading(s) from the required textbook, review the lecture notes, watch the supplemental lecture videos, complete quizzes, programming assignments, and individual projects, and participate in the discussion board. There is also a final group project. Students are expected to complete all deliverables by their assigned due dates.

Course Communication

Canvas will be used for posting relevant course materials throughout the term, and it will serve as the primary mechanism for course communication. I will be active on Canvas, typically every weekday. You are encouraged to post your course-related questions on the "Ask Your Instructor" forum on the course discussion board where other students will be able to benefit from your inquiries. To discuss personal or sensitive concerns, please email me directly at evrenh@bc.edu. You can expect me to respond to questions by email within 12 hours. If you do not hear back from me within 24 hours of sending an email, please resend your message.

There are no set office hours. You can use the office hour appointment page to schedule a virtual live meeting. I will be available throughout the week. I encourage you to be proactive in getting questions answered through the many available methods.

Assessments and Grading Policy

The grading distribution is as shown (below), but additional details on the various course requirements are also described on Canvas.

Component	Weight
Quizzes - There are 7 quizzes based on the course material. - The quiz is timed for 2 – 4 hours, so please allow sufficient time to take each quiz without interruption. - I will review the quizzes and release the solutions within 24 hours after the deadline.	**20%**
Programming Assignments - There are 7 programming assignments (R labs). Each lab consists of two parts. - The first part is to follow through the textbook's lab on your own time. - Once you feel comfortable with the R code and material discussed in the textbook's lab, then you complete the second part. - This second graded part will test your knowledge of the R commands and outputs covered in the textbook's lab. - You will have 2 – 3 hours to complete this second part of the lab, so please allow sufficient time to complete without interruption. - R Lab solutions and code will be released within 24 hours after the deadline.	**20%**
Individual Projects - There are 2 individual projects used to re-enforce course concepts and provide implementation experience in R.	**20%**
Discussion Forum Participation - There are 6 discussion forum assignments on Canvas. - In order to receive full credit, you must submit your initial post by Friday and your three responses by Sundays.	**10%**
Final Group Project - Students will conduct a final course project with their Critical Thinking/Study Group (4 – 5 students) using the machine learning techniques covered in class. - Only one project report should be submitted per group.	**30%**
TOTAL	**100%**

The graduate grading system for Woods College is as follows:

A (4.00), A- (3.67)
B+ (3.33), B (3.00)
B- (2.67)
C (2.00)
F (.00)

Letter Grade	Range
A	100% – 94%
A-	93.9% – 90.0%
B+	89.9% – 87.0%
B	86.9% – 84.0%
B-	83.9% – 80.0%
C+	79.9% – 77.0%
C	76.9% – 74.0%
C-	73.9% – 70.0%
D+	69.9% – 67.0%
D	66.9% – 64.0%
D-	63.9% – 62.0%

All students can access final grades through Agora after the grading deadline each semester. Students who complete course evaluations can access grades earlier, as they are posted.

Course Assignments

Student participation in weekly online discussions is an integral part of this course. The goal of the discussions is to interactively discuss course-related content to achieve the desired learning objectives.

- All online discussion participation requirements must be met to earn the participation points forthat week.
- Students can participate day or night, and participants have time to reflect on their response andcreate succinct easy-to-read replies (quality is preferred over quantity).
- It is imperative to remain respectful of all viewpoints and positions and, when necessary, agreeto respectfully disagree.
- While active and frequent participation is encouraged, cluttering the discussion board with

inappropriate, irrelevant, or insignificant material will not earn additional points and may resultin receiving less than full credit. Frequency is not unimportant, but content of the message is paramount.

- Earning full participation credit each week requires a complete response to a video segmentrelated to analytics from industry-leading organizations.
- After viewing the video, you will discuss what you found to be interesting and what you learned,while answering several questions related to your professional experience and the speaker's approach.
- The first post of your question must be done by Friday evening to receive full credit. You should then return on another day later in the week, reading through the discussion threads, and provide three replies to another student's posts, by Sunday evening.
- For the replies, students should demonstrate that they have read through the discussion as it has matured during the week.
- Posting an initial response after Friday (late) or replying with"good point" or other non-substantive replies, earn no participation credit for the week.
- *Please remember to cite all sources (when relevant) in order to avoid plagiarism.*

Quizzes, programming assignments, and individual projects using R will be used to reinforce both machine learning concepts and practice. The final group project will involve students applying multiple machine learning methods to solve a practical business problem. You can expect me to provide feedback, grade and return assignments within 10 days. Unless otherwise noted, all work is due on the assigned day by 11:59 PM (Eastern Time). Please do not hesitate to ask if you have questions or concerns.

It is expected that you will spend 12 hours per week on out-of-class assignments and exercises. Please note that some weeks will require more time and some weeks less time but the average is approximately 10 hours per week over the term. Materials of the week will be available at the beginning of each week.

Course Schedule

Week / Dates	Topic	Reading	Key Tasks
Week #1 May 18 - May 22	Introduction to Machine Learning	Ch. 1, Ch. 2	- Discussion #1 Due - Quiz 1 / R Lab 1 Due
Week #2 May 23 - May 29	Linear and Nonlinear Regression	Ch. 3, Ch. 7	- Discussion #2 Due - Quiz 2 / R Lab 2 Due
Week #3 May 30 - June 5	Resampling Methods in Machine Learning	Ch. 5	- Discussion #3 Due - Quiz 3 / R Lab 3 Due - Project #1 Assigned
Week #4 June 6 - June 12	Linear Model Selection and Regularization	Ch. 6	- Discussion #4 Due - Quiz 4 / R Lab 4 Due

Week #5 June 13 - June 19	Classification Models	Ch. 4 Handouts	- Discussion #5 Due - Quiz 5 / R Lab 5 Due - Project #2 Assigned
Week #6 June 20 - June 26	Tree-Based Methods	Ch. 8	- Discussion #6 Due - Quiz 6 / R Lab 6 Due
Week #7 June 27-July 1	Support Vector Machines Unsupervised Learning(Optional)	Ch. 9, Ch. 10 (Optional)	- Quiz 7 / R Lab 7 Due - Final Group Project Due

Participation/Late Policy

Students are expected to participate in and complete all discussions, assignments, and assessments. Late assignments will be accepted for up to 2 days with a 10% penalty for each day that assignment is late unless the student has arranged an extension ahead of time with me (I am flexible with everyone's challenging circumstances and time constraints), with rare exceptions based on individual cases (e.g., inability to communicate with me ahead of time because of an emergency).

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.

Accommodation and Accessibility

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 is 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.

Scholarship and Academic Integrity

Students in Woods College courses must produce original work and cite references appropriately. Failure to cite references is plagiarism. Academic dishonesty includes, but is not necessarily limited to, plagiarism, fabrication, facilitating academic dishonesty, cheating on exams or assignments, or submitting the same material or substantially similar material to meet the requirements of more than one course without seeking permission of all instructors concerned. Scholastic misconduct may also

involve, but is not necessarily limited to, acts that violate the rights of other students, such as depriving another student of course materials or interfering with another student's work. Please see the Boston College policy on academic integrity for more information.

Health Integrity Policy

Particularly during this time of the COVID-19 pandemic, we must take even greater measures to care for ourselves, for each other and for our community. Therefore, we are asking that all students care of themselves by monitoring their health and washing their hands thoroughly before class. Feel free to wear a mask, if you wish. We ask that you keep an appropriate physical distance between yourself and others and to not attend class if feeling unwell.