Statistical Machine Learning (SML) Methods for Applied Research/Applied SML
ABE 6933 Section ASML/ASMO, 3 credit hours
Class Periods: TR periods 4-5 (10:40 AM - 12:35 PM)
Location: online: sync (ASML) or async (ASMO)
Academic Term: Fall 2023

Instructor:
Dr. Nikolay Bliznyuk
Email: nbliznyuk [AT] ufl.edu
Phone: 352-392-1864 (only by prior appointment)
Office: 239 Rogers Hall; office hours by Zoom, times TBA

Teaching Assistant/Peer Mentor/Supervised Teaching Student:
- Ms. Madison Moore, moore.madison [AT] ufl.edu

Course Description
Tools of statistical machine learning for applied research targeted at graduate students in engineering and agricultural and life sciences. Statistical approaches to machine learning are emphasized. Application of statistical ML methods and interpretation of the results rather than formal derivations or programming language implementation of the procedures are prioritized. Unlike STA 6703 (Statistical Machine Learning), this course emphasizes breadth (applications and interpretation of results) rather than depth, and is targeted at less quantitative audiences. This course will be delivered in online mode (synchronous or asynchronous).
If you have a schedule conflict with one of Tuesday or Thursday slots, enroll in the ASMO section (async). Recordings from the ASML section (sync) will be made available to all enrolled students (sync and async).

Course Pre-Requisites / Co-Requisites
Prerequisites: a prior course in applied linear regression (e.g., STA6093 or STA6166) or permission of the instructor; familiarity with R software.

Course Objectives
- Learn the language of and the principles behind predictive modeling and model validation
- Learn and be able to use R to apply different classes of statistical learning methods and interpret results, thereby expanding the arsenal of available analytical tools of researchers familiar with regression
- Establish command of methods, their application and reporting of the results through homework, exams and a final project
- Reinforce the use of R as a statistical computing language for data science – for statistical inference, prediction, scientific computing and data visualization

Materials and Supply Fees
None

Required Textbooks and Software
- *An introduction to statistical learning*
- James, G., Witten, D., Hastie, T., & Tibshirani, R.
- 2017, 1st ed, 7th printing or later, Springer
- ISBN 1461471370

- *An introduction to statistical learning*
- James, G., Witten, D., Hastie, T., & Tibshirani, R.
- 2021, 2st ed, latest printing (see www.statlearning.com), Springer
- ISBN 1071614177

R language and R Studio environment are freely available

Recommended Materials
N/A

Course Schedule (tentative)

Week	Topics
1	Course logistics; introduction to SML; nearest neighbors
2	Calculus-based probability: essential review of rvs, independence, moments, etc
3	Mathematical statistics: essential review of estimation and hypothesis testing
4	Maximum likelihood estimation (MLE) and its connections with Bayesian inference
5	Linear regression essentials
6	Linear regression in matrix form and its extensions
7	Classification: logistic regression; discriminant analysis and its flavors
8	Out-of-sample performance metrics: cross-validation and ROC curves
	Midterm
9	Model selection and regularization; penalized regression (ridge, lasso)
10	Spline-based models, generalized additive models
11	Classification and decision trees (single)
12	Ensembles of trees: bagging, random forest, gradient boosting
13	Support Vector Machines
14	Unsupervised learning: dimension reduction and clustering
15	Advanced topics (high-dimensional inference, multiple testing, deep learning, etc)
	Project report OR take-home final exam due

Depending on the typical student background, the above schedule may be adjusted to reflect the interests of the class (e.g., less time on the probability and statistics review and more time on advanced topics).
In particular, in Fall 2023, we'll likely expedite the (sync lecture) coverage of topics from weeks 2,3,6 so that we can have more time for more advanced topics. Full coverage of these topics (2,3,6) will be provided through lecture recordings.

Attendance Policy, Class Expectations, and Make-Up Policy
Attendance at all times is expected. Make-ups are not allowed except for documented health, family emergency or work reasons. Excused absences must be consistent with university policies in the Graduate Catalog (https://catalog.ufl.edu/graduate/regulations) and require appropriate documentation. Additional information can be found here: https://gradcatalog.ufl.edu/graduate/regulations/

Evaluation of Grades (Tentative)

Assignment	Total Points	Percentage of Final Grade
Homework Sets (8)	100 each	25%
Quizzes (8)	100 each	25%
Midterm	100	25%
Final Exam or Project	100	25%
		100%

Quizzes (online, in Canvas) and homework assignments will be closely matched to the course topics and will span approximately 4 hours of lectures. All quizzes and all homework sets will be weighted equally by converting the score to the 100-point scale first. Extra-credit opportunities to gain "bonus points" will be provided periodically. Project details are provided at the end of the syllabus. Tentatively, the option of course project will only be available if the enrollment is not excessively high to allow it to be graded in time before the university-mandated deadline.

Grading Policy

Tentatively, the following grading scale will be adopted; grades may be curved to the advantage of students.

Percent	Grade	Grade Points
93.4 - 100	A	4.00
90.0 - 93.3	A-	3.67
86.7 - 89.9	B+	3.33
83.4 - 86.6	B	3.00
80.0 - 83.3	B-	2.67
76.7 - 79.9	C+	2.33
73.4 - 76.6	C	2.00
70.0 - 73.3	C-	1.67
66.7 - 69.9	D+	1.33
63.4 - 66.6	D	1.00
60.0 - 63.3	D-	0.67
0 - 59.9	E	0.00

More information on UF grading policy may be found at:
UF Graduate Catalog
Grades and Grading Policies

Students Requiring Accommodations
Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation
Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

In-Class Recording
Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by

a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.

University Honesty Policy

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment

The Herbert Wertheim College of Engineering values varied perspectives and lived experiences within our community and is committed to supporting the University's core values, including the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of race, creed, color, religion, age, disability, sex, sexual orientation, gender identity and expression, marital status, national origin, political opinions or affiliations, genetic information, and veteran status.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:
• Your academic advisor or Graduate Program Coordinator
• HWCOE Human Resources, 352-392-0904, student-support-hr@eng.ufl.edu
• Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.ufl.edu
• Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: https://registrar.ufl.edu/ferpa.html

Campus Resources:

Health and Wellness

U Matter, We Care:
Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: https://counseling.ufl.edu, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence
If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the **Office of Title IX Compliance**, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)
Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or http://www.police.ufl.edu/.

Academic Resources

E-learning technical suppor*t*, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. https://lss.at.ufl.edu/help.shtml.

Career Connections Center, Reitz Union, 392-1601. Career assistance and counseling; https://career.ufl.edu.

Library Support, http://cms.uflib.ufl.edu/ask. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring. https://teachingcenter.ufl.edu/.

Writing Studio, 302 Tigert Hall, 846-1138. Help brainstorming, formatting, and writing papers. https://writing.ufl.edu/writing-studio/.

Student Complaints Campus*:* https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/;https://care.dso.ufl.edu.

On-Line Students Complaints*: https://distance.ufl.edu/getting-help/; https://distance.ufl.edu/state-authorization-status/#student-complaint*

Supplements:
Project description (tentative, will be finalized in class)
The project will emphasize creative application of the methods developed in the course. Ideally, the application would be to your line of research and data (your own or of your immediate collaborators - advisor or fellow students). If you do not have suitable data, please check out the sources at the end of this description for the publicly available datasets. Otherwise, a good project could be a replication and extension of the results of a paper of interest that uses the methods from our course. "Creative application" does not allow merely running someone else's code without making other contributions. **Plagiarism is totally inappropriate and prohibited (just do not do it); it will result in a failing grade for the course. Course staff will run all project reports through UF Ithenticate. All work should be done individually (unless explicitly permitted by the instructor – for more ambitious projects).** *Projects already completed for other classes/causes are not acceptable. Example 1 - unacceptable: in a previous semester, a student wrote a paper for a journal or did a project for a different class, and now wants to submit it without major changes or additional SML type of work as the SML class project. Example 2 – acceptable: in a previous semester, a student wrote a paper for a journal or did a project for a different class, but wants to do a major extension of the work using the techniques learned in the SML class. This would make a potentially very good project, but the student needs to be explicit about what is new and what not. Only the new work will constitute the course project in this case.* The project will be used to assess the knowledge and skills that students acquired in the course; for that reason, the work must be done individually and without assistance from the course staff.
Deliverables: a one-page proposal and a short technical report as described below.
Deadlines: (tentative and will be revised appropriately each year)

(TBA; tentatively, mid-November): submit your proposal by email to nbliznyuk@ufl.edu if you are planning to complete the optional project, so that we can meet on Thursday during the class time (individual slots TBA after your proposals have been received).

(TBA; tentatively, early during the exams week): final report (in pdf format, accompanying code and the actual data that you used, if using a publicly available source; put all in a folder named after you and create a zip or rar archive; test archive before submitting), submit using Dropbox file request (same link as above): tinyurl.com/nbliznyuk-submit-files

Expectations for the proposal (1 page):

The main goal behind the proposals is to ensure that the projects are neither too simple nor too ambitious (i.e., will require about 30 hours to complete – loosely, an equivalent of 3-4 weeks of homework effort, where writing will play a significant role), there is no duplication among students and that you have the necessary relatively clean data to analyze. Please specifically discuss what you propose to do (e.g., "big picture" and specific methods), why you focus on this particular problem (significance, motivation and relevance to the course) and available data (specifically, what are primary response variable(s) and features, what are n and p, etc). Your project should be "shovel ready", i.e., a bit of data preprocessing may be necessary but you should not be spending more than 20% (ideally, 10%) of your total time budget on cleaning and data manipulation. The typesetting format of the proposal should be the same as for the project (please read below).

Expectations for the report (8 pages):

Report should be organized as a short paper appropriate to your field; e.g., a short abstract (100 words), intro (including motivation), background and data, methods, analysis/results, conclusions/discussion. *Any software/languages/environments may be used for the project (i.e., not necessarily R). In most cases, the project should use several classes of methods (multiple linear regression or logistic regression as the baseline – possibly coupled with variable selection; at least one shrinkage method (if p is high) and/or a GAM (if p is low), and a tree-based ensemble method, typically, random forest (and definitely discuss variable importance summaries that may be extracted from it) and possibly boosting (because it is often the best method), ideally both) for classification or regression and examine out-of-sample performance of the methods using K-fold cross-validation.* If you are doing binary classification, please additionally include and discuss the ROC curves. If doing non-binary classification, please showcase a binary classification subproblem and include and discuss the ROC curves. The length is about 8 pages (not counting references, appendices or supplements) double-spaced, using 12 pt font: roughly 6 pages of text and 2 pages for your most essential tables and figures; **single-column only**. If necessary, the paper may have an Appendix with additional figures and tables. Data, code and other supplemental information should be made available as part of "Supplementary Materials" unless the data are confidential (please discuss "deliverables" in the proposal). Please check out the project evaluation rubric in a separate file.

Some sources of data for projects:

Google search "data for machine learning", e.g.,

UCI ML Repository

http://archive.ics.uci.edu/ml/

Kaggle Competitions

https://www.kaggle.com/datasets

PLOS One journal website

read info about data availability in papers of interest and at http://journals.plos.org/plosone/s/data-availability