BUAN 5310 Statistical Learning
Winter 2020

1. Course Detail

Professor:	Dr. Misuk Lee
Office:	Pigott 413
Email:	leem@seattleu.edu
Office Hours:	TBD
Website:	All assignments and class materials will be posted on Canvas.

2. Course Objective

Many problems in today's business require traditional and nontraditional forms of data analysis. In particular, rapid developments in data collection and storage technologies have led to big data sets and new questions;

- Amazon collects purchase histories and item ratings from millions of its users. How can it use these to predict which items users are likely to purchase and like?
- Yahoo news acts as a clearinghouse for news stories and collects user click-through data on those stories. How should it organize the stories based on the click-through data and the text of each story?
- How does Netflix recommend movies to each of its users?

An expert's answer to any one of these questions may very well contain enough material to fill its own course, but basic answers stem from the principles of statistical learning and data mining.

Upon completing this course, you should be able to tackle real world data mining problems, by: (1) selecting the appropriate methods and justifying your choices; (2) implementing these methods programmatically (using, say, the Python programming language) and evaluating your results; (3) presenting your results to broad audiences in business outside statistics and/or data mining.

At present, I plan to cover the following topics:

- **Supervised Learning**: multinomial logistic regression, naive Bayes, nearest neighbor classifier, decision tree, support vector machine, neural networks.
- **Unsupervised Learning**: clustering, association analysis, sequence mining, text mining.

The focus of the class will be to interactively discuss contemporary topics of statistical learning and data mining. The methodology will include lecture, in-class exercise, minicase studies, and project.

3. Recommended Texts

- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Pearson, ISBN-10: 0321321367, ISBN-13: 978-0321321367
- Reference
  - Andreas C. Mller and Sarah Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, ISBN-13: 978-1449369415, ISBN-10: 1449369413
  - James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning, Springer, ISBN 978-1-4614-7137-0

4. Assessment

- Two Exams: 60 % (30 % each)
- Homework/Attendance: 10 %
- Group Project: 30%

I will also offer you extra credit for exceptional class participation during lecture.

5. Grade

Final grades will be assigned based on the following rubric:

- 95 to 100: A, 90 to 94 : A-
- 85 to 89 : B+, 80 to 84 : B, 77 to 79 : B-
- 74 to 76 : C+, 70 to 73 : C, 67 to 69 : C-
- 64 to 66 : D+, 60 to 63 : D, 50 to 59 : D-

6. Homework

Homework will be posted on Canvas. You are allowed to discuss class exercise problems with your fellow students. However, the work you submit must be your own. You must acknowledge in your submission any help received on your assignments. That is, you must include a comment in your submission that clearly states the name of the student, book, or online reference from which you received assistance.

7. Group Project

Through the project, you will gain experience in a process of data mining. This project is an integral part of the course, since it allows students to apply the concepts, methodologies, and tools in the context of a real-world application. You will select a real-world application and perform data mining technique(s) on the real dataset using Python and present an analysis report on (1) goal of project, (2) description of the quality/features of data (data processing), (3) data mining techniques adopted, (4) performance evaluation, and (5) final findings or conclusions. More details about the final project will be announced in class.

8. Non-Disclosure Agreement

Data, codes, and lecture materials, including assignments, projects, lecture notes, and others, should not be shared with any person or organization outside of class. If you have a special reason to share any of the materials with someone else, please ask for permission first.

9. Course Policy

- If you miss a test due to health issues for self or family and work-related emergency, you will need to provide appropriate documentation.
- All assignments are to be turned in through Canvas.
- The use of computers during class is only limited to course-related purposes. If a student misuses a computer (such as web browsing, chatting, doing homework, etc), it will lower his/her final grade (10 percent off for each incident from your homework grade).

10. Regrading

Regrade requests must be made in writing and attached to the exam. The request must include a description of your objection and why you think your exam should have been graded differently. If a test is submitted for regrading, I will regrade the entire test- so it is possible to either gain or lose points. This policy does not apply to arithmetic errors.

11. Disability

If you have, or think you may have, a disability (including an invisible disability such as a learning disability, a chronic health problem, or a mental health condition) that interferes with your performance as a student in this class, you are encouraged to arrange support services and/or accommodations through Disabilities Services staff located in Loyola 100, (206) 296-5740. Disability-based adjustments to course expectations can be arranged only through this process.

12. Notice Regarding Religious Accommodations

It is the policy of Seattle University to reasonably accommodate students who, due to the observance of religious holidays, expect to be absent or endure a significant hardship during certain days of their academic course or program. Please see, Policy on Religious Accommodations for Students (https://www.seattleu.edu/media/policies/Policy-on-Religious-Accommodations-for-Students—FINAL.PDF).

13. Honor Code

Seattle University is committed to the principle that academic honesty and integrity are important values in the educational process. Academic dishonesty in any form is a serious offense against the academic community. Acts of academic dishonesty will be addressed according to the Seattle University Academic integrity Policy.

Academic Policies on Registrar website (https://www.seattleu.edu/redhawk-axis/academic-policies/)

- Academic Integrity Policy
- Academic Grading Grievance Policy
- Professional Conduct Policy (only for those professional programs to which it applies)

If you are not sure whether a particular action is acceptable according to the Academic Integrity Policy, you should check with your instructor before engaging in it.

14. Office of Institutional Equity

Title IX of the Education Amendments of 1972 (Title IX) prohibits discrimination based on sex in educational programs or activities that receive Federal financial assistance. This prohibition includes sexual misconduct, which encompasses sexual harassment and sexual violence. Seattle U remains committed to providing a safe and equitable learning, living, and working environment. Seattle U offers emergency, medical, and other support resources, as well as assistance with safety and support measures, to community members who have experienced or been impacted by sexual misconduct.

Seattle U requires all faculty and staff to notify the University's Title IX Coordinator if they become aware of any incident of sexual misconduct experienced by a student. For more information, please visit https://www.seattleu.edu/equity/. If you have any questions or concerns, you may also directly contact the Title IX Coordinator in the Office of Institutional Equity (email: oie@seattleu.edu; phone: 206.296.2824) University Resources and Policies.

15. Academic Resources

- Library and Learning Commons (http://www.seattleu.edu/learningcommons/) (This includes: Learning Assistance Programs, Research [Library] Services, Writing Center, Math Lab)

- Academic Integrity Tutorial (found on Canvas and SU Online)

16. Tentative Course Schedule (Attached Below)

16. Tentative Course Schedule

Week	Topic	Note
1 (1/9)	Overview & Logit, Multinomial Logit	
2 (1/16)	Naive Bayes, Nearest Neighbor Classifier	
3 (1/23)	Decision Tree	
4 (1/30)	Neural Networks, Support Vector Machine	
5 (2/6)	Exam	Exam 1
6 (2/13)	Model Evaluation, Clustering	
7 (2/20)	Association Analysis	
8 (2/27)	Text Mining, Sequence Mining	
9 (3/5)	One-on-one meeting	
10 (3/12)	Ensemble, Bagging, Boosting, Anomaly Detection	Group Project Due
11 (3/19)	Exam	Exam 2

Course schedule is subject to change without announcement.