BUAN 4310 Data Mining and Big Data

Albers School of Business and Economics
Seattle University

1. General information

Instructor:	Ben Yeo (Master Yeoda)
Email:	byeo@seattleu.edu
Classroom:	NA
Day/Time:	Online
Office hours:	Tue & Thu; 1230 – 1330 hrs
Pre-requisite:	BUAN 3210

2. Course description

Business analytics involves the use of data driven decision-making techniques. Among which is data mining. Businesses, governments, and non-profit organisations collect and have access to large amounts of data about customers, suppliers, employees, regions, population, transactions, and such. Data mining covers a wide range of techniques to detect actionable patterns in data and generate predictions. In the course, we will work with various data problems. We will examine types of questions that data mining can answer and the appropriate data mining techniques for answering different questions. The emphasis is on understanding the concepts and logic behind a wide range of data mining techniques and their relation to business analytics. By completing the course, students will be better equipped to apply these techniques to address various business problems.

3. Faculty biographical sketch

Yeoda is a faculty at Seattle University. His research interests include innovation- and technology-driven economic growth and social informatics, using both quantitative and qualitative methods, including traditional statistics and data mining. Prior to academia, Yeoda worked in the industry as a Senior Research Analyst in Economics. He received a Ph.D. in Information Science from the College of Information Sciences and Technology at the Pennsylvania State University. Obviously, a Star Wars nerd, he is ☺

4. Learning objectives

The primary course outcome is to provide students with an understanding of data mining techniques and apply them to solve different business problems. Specific learning objectives include:

1. Become familiar with the applications of data mining in today's data rich-environment
2. Gain experience using key data mining methods of classification, prediction and exploration
3. Know how to decide when to use which technique
4. Be able to implement major data mining techniques using software

5. Gain the intellectual capital required to provide responsible business analytics services
6. Define key terminology in model fitting and statistical learning
7. Pre process data for data mining
8. Compute and interpret descriptive statistics and graphical techniques
9. Build, analyse and interpret results of various statistical learning techniques

5. Course materials

In this course, we will use the R programming environment and language. R is a highly flexible tool, and it is now ubiquitous in modelling across many disciplines. R is not a point and click application like SPSS or Stata; it requires some programming skills, which students should already have a basic understanding. This course uses RStudio the environment for the language. Since the course focuses on the techniques and applications, rather than the R syntax, students may use XLMiner (with a graphical user interface) if they prefer. Access to XLMiner will be discussed in class. Additional materials may be assigned as needed.

- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). Data mining for business analytics: Concepts, techniques, and applications in R. John Wiley & Sons.

6. Evidence students will submit

Grades are based on course deliverables. Specifically, there are 4 graded components for the course. These are given as follows.

- Recaps x6
- Labs x4
- Term Project
- Mid Term
- Final exam

Recaps are quizzes to be taken on Canvas. These are both conceptual and applied, and geared towards keeping us on track.

Labs are exercises where students apply the concepts to solve small business problems using data solutions. In an on-ground setting, these would be class activities. In an online format, these are to be completed and submitted on Canvas.

The term project is problem-based, where students apply the tools covered to develop solutions to solve the given problem(s). The project comprises 2 large data problems leading to a presentation[1] and a final report[2]. Students should complete these projects in groups of 2-4

---

[1] This is to be submitted as a deck of annotated power point slides or a video recording of the presentation. Given the online format, no in-person presentation will be done.
[2] All project deliverables cannot be graded twice. Students can submit multiple times before the corresponding deadlines, but only the latest submission before the corresponding deadlines will be taken into account.

members[3] each. All deliverables should be submitted on Canvas. Feedback will be provided to submitted work on Canvas approximately within one week of submission. **All late submissions of any deliverable without an approved extension[4] will receive a 5% penalty for each day late**. The quality of submissions will be measured using corresponding rubrics, available on the assignment itself. However, some aspects may involve subjective evaluation. As much as possible, the rubrics will frame the assessments5. All deliverables will be evaluated on evidence of learning, depth of analysis, organisation and thoroughness. They should demonstrate the consideration of key issues and critical thinking with respect to the data mining algorithms applied in a business setting. All written deliverables should be professional prepared, typed, and contain a bibliography of cited sources and appropriate footnotes (where applicable). All ideas, quotes and statistics borrowed from another author/source must be cited. If external content (such as facts or data) are used, please be sure to cite the corresponding authorities. Students may use any citation style, formal or otherwise. The objective here is simply to attribute to corresponding resources accordingly.

The mid term comprises a series of short conceptual, factual and some applied questions that students should be able to address within 90 minutes. There are likely to be approximately 60[6] questions. **Make ups must be completed within a week of the original test date.** However, this requires appropriate documentation, such as medical documentation, for a valid and/or official reason to justify the absence. **Make ups will be different (i.e. different questions) and thus, may not be at the same level of difficulty**. If a student requires special accommodation, he/she must inform the instructor in advance with the appropriate documentation so special arrangements can be made.

The final exam follows a similar format to the mid term, except it is cumulative[7]. Hence, it is weighted more heavily. Make ups for the final exam require sufficient documentation (such as medical), and must be taken within the same finals week. **Make ups for the final exam require sufficient documentation (such as medical), and must be taken within the same finals week.** Per the University's policy, INCOMPLETE grades can only be given un extenuating circumstances. This does not include missing the deadline for a final project because of connection issues, or having a common cold on the day of the exam. Make ups will be different and may not be at the same level of difficulty. Likewise, if a student requires special accommodation, he/she must inform the instructor in advance with the appropriate documentation so special arrangements can be made.

---

[3] A peer evaluation of each group member's contributions is available at the end of the term. Students may choose not to submit a peer evaluation if they think everyone contributed equally. This score will affect the final project score.

[4] Extensions on any deliverable have to be very well justified on reasonably valid reasons and with appropriate documentation. The deliverables are assigned well in advance and any delays should be anticipated. Last minute extensions are not acceptable except under emergency situations.

[5] All grade disputes (if any) must be addressed immediately. No grade changes to the deliverables are allowed 1 week after the grade for the corresponding deliverable is released.

[6] -/+ 10%. Subject to minor changes as needed. The format will be clearly communicated in class prior to the test date.

[7] Several concepts are related and build upon each other.

6.1 Expectations

Participation and studying outside class times are critical in completing the course. Students are expected to read the assigned materials according to the course schedule.

Participation is critical in completing the course. Absence from more than two weeks can lead to a failing grade[8]. Students should contact the instructor if he/she anticipates missing a class. Students are responsible for reviewing materials posted on Canvas. Students are expected to be present in all classes barring unforeseen circumstances and official/professionally-related absences due to necessity. In addition, students are expected to participate in class discussions to facilitate a healthy exchange of ideas and an effective learning experience for everyone. Class participation comprises (i) class attendance and (ii) substantive contributions to class discussions.

6.2 Criteria for assessment

The final grade will be based on the graded deliverables

Recaps (x6):	5%
Labs (x4):	5%
Project:	48%[9]
• Problem 1 22%	
• Problem 2 26%	
Mid Term:	20%
Final exam:	22%
Total:	**100%**

6.3 Grade descriptions

Grading will follow the Albers School's recommended rigour. Students are expected to follow the rubrics very closely.

Grade	Point	Grade Range	Grade Scale
A	4.0	95% +	The instructor judged the student to have accomplished the stated objectives of the course in an OUTSTANDING manner.
A-	3.7	90 – 94.99%	
B+	3.3	87 – 89.99%	
B	3.0	83 – 86.99%	The instructor judged the student to have accomplished the stated objectives of the course in an ACCEPTABLE manner.
B-	2.7	80 – 82.99%	
C+	2.3	77 – 79.99%	
C	2.0	73 – 76.99%	The instructor judged the student to have accomplished the stated objectives of the course in a POOR manner; but it is still a passing grade.

---

[8] This will be approximately 20% of the course, which is synonymous with missing a day's work in a week.

[9] The final score may be modified by peer evaluations where applicable.

Grade	Point	Grade Range	Grade Scale
C-	1.7	70 – 72.99%	
D	1.0	60 – 69.99%	The instructor judged the student to have accomplished the stated objectives of the course in a DREADFUL manner. (A grade of D will not fulfil the requirements in a major field of concentration.)
F	0.0	Less than 60%	The instructor judged the student NOT to have accomplished the stated objectives of the course. TROLL

7. Course schedule[10]

Classes will comprise a mixture of lectures, demos, and labs. Lectures are geared towards explaining the concepts behind the techniques. Demos illustrate how the techniques work. Labs are hands-on activities where students will solve a business problem in class. These are designed to provide students with sufficient knowledge and experience with these techniques. In addition, homework is assigned on most weeks to provide conceptual reviews and hands-on practice.

Week Date	Topic	Activities[11]
Week 1 Sep 9 – 11	Introduction to Data Mining	Readings: • Galit et al. Ch 2 Activities: • Course overview • Conceptual overview • Getting into project groups[12] • Lab: Software installation Homework: • Recap 1 Intro to Data Mining (Due Sep 13)

---

[10] Class schedules are tentative and are subject to changes as necessary.
[11] Additional readings will be assigned as needed.
[12] For those who are unable to find a group of 2-4 students in time, the instructor will randomly assign them.

Week Date	Topic	Activities[11]
Week 2 Sep 14 - 18	Data Visualisations and Exploration	Readings: • Galit et al. Ch 3 Activities: • Demo: Alita's World Time Heist Homework: • Recap 2 Data Visualisation (due Sep 20) • Term project assigned (check due dates for different parts)[13] o Problem 1 is due on Nov 1 o Problem 2 is due on Nov 15
Week 3 & 4 Sep 21 – 25; Sep 28 – Oct 2	K Nearest Neighbours	Readings: • Galit et al. Ch 7 Activities: • Demo: k Nearest Avengers • Lab 1: Bank Loans Homework: • Lab 1: Bank Loans (Due Oct 4) • Recap 3 kNN (Due Oct 4)
Week 5 & 6 Oct 5 – 9; Oct 12 – Oct 16	Mid Term and Decision Trees	Readings: • Galit et al. Ch 9 Activities: • **Mid Term** (**Taken on Oct 6 on Canvas**) • Demo 1: A Padawan in Hogwarts • Demo 2: Dungeons and Dragons • Lab 2: Used Car Sales Homework: • Lab 2: Used Car Sales (Due Oct 18) • Recap 4 Decision Trees (Due Oct 18)

---

[13] Due to the nature of the materials, there is much to cover before a project can be completed. Hence, the due dates are in the second half of the course. Nonetheless, project groups should start working on the project in Week 2 and ask questions (if any) along the way.

Week Date	Topic	Activities[11]

Week 7 Oct 19 – Oct 23	Unbalanced Data; Model Diagnostics	Readings:
		• NA
		
		Activities:
		• Unbalanced data:
		  o Demo: Every Rose has Its Thorn
		• Model diagnostics:
		  o Demo: Diagnostics of the Hangover

Week 9 Oct 26 – Oct 30; Nov 2 – Nov 6	Regression	Readings: • Galit et al. Ch 6 Activities: • Demo: Glory! Glory! Man United! • Lab 3: Airfare prediction Homework: • Lab 3: Airfare prediction (Due Nov 8) • Recap 5 Regression (Due Nov 8)
Week 10 & 11 Nov 9 – 13; Nov 16 – 18	Hierarchical Clustering & Q&A	Readings: • Galit et al. Ch 15 (Sections 15.1 – 15.4) Activities: • Demo: Clustering the Star Wars Universe • Lab 4: Universities • Q&A Homework: • Lab 4 Universities (Due Nov 15) • Recap 6 Clustering (Due Nov 15)
Finals Nov 19 - 24	Final exam week	Check SU final exam schedule[14]

8. Academic resources

• Library and Learning Commons (http://www.seattleu.edu/learningcommons/)

[14] Please note the final exam schedule. **Family/personal travel plans (and vacation) are not acceptable reasons for missing the final exam.** There is sufficient time to plan in advance.

- This includes: Learning Assistance Programs, Research [Library] Services, Writing Center, Math Lab
- Academic Integrity Tutorial (found on Canvas and SU Online)

9. Academic policies

Various Seattle University academic policies can be found at the following URL. Please be sure to review them prior to the course.

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)

10. Notice for students concerning disabilities

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.

11. Notice on 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-onReligious-Accommodations-for-Students---FINAL.PDF

12. 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