

Course Information

Course Title: Data Mining and Predictive Analytics	Course Number: BUDT 758T
Meeting Times: Tuesdays, Thursdays 9:30-10:45am (Section 0501) 12:30-1:45pm (Section 0502) 2:00-3:15pm (Section 0503)	Location: TBA
Term: Spring 2022	Credits: 3
Professor: Dr. Jessica M Clark	Pronouns: She/Her/Hers
Email: jmclark@umd.edu Cell Phone: (978) 870-9654 Office Location: 4308 Van Munching Hall (Decisions, Operations, and Information Technologies) Office Hours: Tuesdays and Thursdays 3:30-4:45pm, by appointment using the Google Calendar appointment page posted on Canvas. Please email to set another time if this is not convenient for you.	
TA: TBD	
Course Dates: Course runs from January 24 2022 – May 12 2022	

Course Description

In business magazines, on TV, and in boardrooms, “big data” and “data analytics” are hot topics. Vast quantities of data are being generated, including new types of data such as web traffic, social network data, and reviews and comments on websites. This data is a valuable resource that, when used correctly, can create a competitive edge for companies by improving the quality of decision making. Recent advances in computing hardware and software have made the application of advanced analytical methods much easier. This course takes advantage of these developments to introduce data analytics to those interested in developing expertise in data-driven decision making.

The course is intended to provide an introduction to the tools and techniques of data mining & machine learning that are central to business analytics, with particular emphasis on classification and prediction. The focus will be on business applications and examples from Marketing, Finance, Healthcare, and Operations will be used to illustrate the breadth of applications.

Course Objectives

Students can expect to:

- Learn fundamental data analytic concepts and techniques required to make effective use of large amount of data
- Develop an understanding of important ideas in regression, classification, association and causation; and learn how these can aid better decision-making

- Study a range of business applications – such as targeting in marketing campaigns, customer retention, fraud detection, loan default prediction, collaborative filtering – so as to be better equipped to develop effective business solutions
- Learn how to avoid common pitfalls and ensure a successful data-mining project outcome
- Gain hands-on experience using realistic datasets and data-mining software

Time Expectations

This course will consist of videos, synchronous discussions and hands-on lab sessions. Research has shown that individuals learn better by interacting and engaging with the material, and so the course is highly interactive. Prepare to practice R programming, complete math exercises, or do some other interactive activity during every session and engage with peers during class time.

Detailed preparation instructions will be provided before each class session, so check the course site before every class. Those instructions supersede the course schedule in this syllabus. A sample work schedule is as follows:

Work Schedule

- **Class Time:** In-class activities and explanations
- **Outside of Class:** Complete assigned readings, assigned videos, and quizzes
- **Outside of Class:** Work on major assignments
- **Outside of Class:** Monitor email/ELMS for updates

Course Materials & Software

1. **Book (Required):** *An Introduction to Statistical Learning with Applications in R* by James, Witten, Hastie, and Tibshirani [ISLR]
 - This book is available online in its entirety with videos and additional help at <http://www-bcf.usc.edu/~garth/ISL/>
2. **Book (Required):** *Data Mining: Practical Machine Learning Tools and Techniques* by Witten and Frank, make sure to get the second edition
 - Free PDF of this book is on the ELMS site
3. **DataCamp (Required):** I will occasionally assign videos from DataCamp to supplement the readings. Sign up for free academic access here:
 - https://www.datacamp.com/groups/shared_links/5d3624a2748d642c6cf22537ed60fc262fee6255809dd27cd478b84a253d60aa
4. **Book (Highly Recommended):** *Data Science for Business* by Foster Provost and Tom Fawcett
 - This is my favorite data mining book. Unfortunately, it doesn't teach any software! You can get an electronic copy on Amazon for about \$20.

Software Requirements

Software Name	How to Access it
R	<ol style="list-style-type: none"> 1. Download R and RStudio 2. Install the tidyverse package 3. You may have to install other packages over the course of the semester

Grading Structure

Your course grade comprises several different components. Note that my grading distribution is not fixed in advance.

Assignment	Percentage %
Homework	15
Quizzes	10
Group Project	20
Midterm	20
Final Exam	25
Classwork and Participation	10
<i>Total</i>	100%

Individual Assignments

- There will be four homework assignments for you to complete individually.
- All assignments (H1-H4) are distributed and submitted through ELMS.
- Feel free to ask questions but only clarification questions will be answered.

Quizzes

- All quizzes are distributed through ELMS.
- Quizzes are worth 10 points each.
- You will have unlimited attempts to complete the quizzes. The purpose of these quizzes is to make sure you understand the important points in the asynchronous reading/videos.
- The quizzes will also serve as good practice for the midterm and final exams.

Midterm and Final

- Exams will be open-book, open notes and given during class time via ELMS.
- Details will be distributed later in the semester.
- The final exam is cumulative and will cover all the material for the semester.

Team Project

- The project this semester will be a Kaggle-style data mining team contest. This project format has been really fun for myself and the teams in the past 😊
- Each team of 3-5 students will work on a prediction task using real data. Everyone will compete to have the most accurate predictions on a held-out dataset. The winning team will receive a prize.
- More details on the final project will be distributed later in the semester.

Professionalism

- **Classwork:** The class sessions will be very hands-on and interactive. To make sure that you're getting as much as possible out of the sessions, you will be expected to turn in a deliverable via ELMS at the end of every class meeting. These will count towards your classwork and participation score.
- **Participation:** I will consider the quality and frequency of your contributions to the class: e.g., producing meaningful, insightful, relevant and concise contributions; helping and encouraging others; actively listening to others' presentations.
- If you have concerns, please contact me as soon as possible, preferably in the first week of the semester.

Course Outline

Not all topics can be covered thoroughly in the discrete time blocks, so the class schedule is flexible and subject to change. The preparation, readings, and assignment due dates will always be announced in class and on Canvas. Check Canvas for the most up-to-date course schedule.

Week	Date	Topic	Tuesday	Thursday	Due**
1	Jan 25, 27	Introduction	Intro to data mining	R Programming Lab	
2	Feb 1, 3	Modeling in R	Regression review	Linear reg in R	Q1
3	Feb 8, 11	Classification	Intro to classification	Logistic Regression	Q2
4	Feb 15, 17	Model Evaluation	Model Evaluation	More Evaluation	Q3, H1
5	Feb 22, 24	Tree-based models	Trees - Part 1	Trees – Part 2	Q4
6	Mar 1, 3	“Black box” models	kNN	Ensemble methods	Q5
7	Mar 8, 10	Putting it all together	Data and features	Predictive systems	Q6, H2
8	Mar 15, 17	Midterm	Exam review	Midterm (in class)*	T1
SPRING BREAK 😊					
9	Mar 29, 31	Complexity	ROC & Lift	Complexity Control	Q7
10	Apr 5, 7	Probabilistic Modeling	Naïve Bayes	More probabilistic models	Q8
11	Apr 12, 14	Text Mining	Text 1	Text 2	Q9
12	Apr 19, 21	Project consultations	Project consultations	Project consultations	H3
13	Apr 26, 28	Unsupervised learning	Association Rules	Cluster Analysis	Q10
14	May 3, 5	Catch-up week	TBD	TBD	H4
15	May 10, 12	Wrap-up	Meet as necessary	Wrap-up and review	Project
		Final Exam*, details TBD			T2

* Major Scheduled Graded Event

** All assignments are due by 11:55pm on the Sunday of that week. H =Homework, Q = Quiz, T=Test.

For Major Scheduled Graded Events (MSGEs), only University approved absences will be accepted (more details below). Make-up exams are at my discretion, and no extra time will be provided for late arrivals for exams.

Course Policies

Professionalism

All students are expected to act with professionalism during all class sessions and interactions with the professor, other students, and any guests. Professionalism covers your conduct in all class activities, homework review, and all class-related communications (e.g., emails with peers and instructor). All emails should be written with a subject and an introduction (e.g., "Dear Dr. Clark"). Use proper spelling, correct grammar, and business-appropriate language (e.g., no slang or inappropriate emojis).

Class Discussions and Participation Guidelines

The primary goal of our discussions is to enhance the collective learning of the class. The following guidelines can help you make effective contributions to our class discussions. Please see the following list of tips for effective and meaningful discussion participation.

Examples of Good Discussion Contributions:

- Comments/questions that provide or seek clarification
- Comments that make points clearly
- Comments that move learning forward by building an argument and/or drawing on other comments without repeating
- Comments that respectfully articulate a point of agreement/disagreement

Examples of Ineffective Discussion Contributions:

- Destructive attacks
- Interrupting peers
- Monopolization of the discussion
- Disrespectfully articulating a point of agreement/disagreement
- Out-of-sequence comments / destroying the flow of class discussion

Office Hours

I will hold regular office hours via Zoom on Tuesdays and Thursdays from 3:30-4:45pm, plus additional hours on an ad-hoc basis. Check my availability and make appointments using this appointment page link:

<https://calendar.google.com/calendar/u/0/selfsched?sstoken=UUFzYUN0TXI1eUI3fGRIZmF1bHR8NjU4MDM4OWI1MjRiRmNmZDkzMzRkZTJjZDdmYzJmMjk>.

If those times are not convenient for you, please email to set an alternative meeting time. You are also welcome to meet with me in small groups.

I would especially appreciate if you could schedule your office hours meetings either immediately before or after appointments that have already been scheduled. This will allow me to use the unscheduled time to be as productive as possible.

Office hours serve several different purposes in addition to answering questions from class and clarifying information. This time provides a way for you to get to know me, and for me to know you, and an opportunity to discuss academic and/or career goals. Note that I **want** you to use office hours for these purposes. I am happy when students attend my office hours!

Resources & Accommodations

Accessibility and Disability Services

The University of Maryland is committed to creating and maintaining a welcoming and inclusive educational, working, and living environment for people of all abilities. The University of Maryland is also committed to the principle that no qualified individual with a disability shall, on the basis of disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of the University, or be subjected to discrimination. The [Accessibility & Disability Service \(ADS\)](#) provides reasonable accommodations to qualified individuals to provide equal access to services, programs and activities. ADS cannot assist retroactively, so it is generally best to request accommodations several weeks before the semester begins or as soon as a disability becomes known. Any student who needs accommodations should contact me as soon as possible so that I have sufficient time to make arrangements.

For assistance in obtaining an accommodation, contact Accessibility and Disability Service at **301-314-7682**, or adsfrontdesk@umd.edu. Information about [sharing your accommodations with instructors, note taking assistance](#) and more is available from the [Counseling Center](#).

Student Resources and Services

Taking personal responsibility for you own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit tutoring.umd.edu to learn more about the wide range of campus resources available to you.

In particular, everyone can use some help sharpen their communication skills (and improving their grade) by visiting ter.ps/writing and schedule an appointment with the campus Writing Center.

You should also know there are a wide range of resources to support you with whatever you might need (see go.umd.edu/assistance). If you feel it would be helpful to have someone to talk to, visit counseling.umd.edu or [one of the many other mental health resources on campus](#).

Basic Needs Security

If you have difficulty affording groceries or accessing sufficient food to eat every day, or lack a safe and stable place to live, please visit go.umd.edu/basic-needs for information about resources the campus offers you and let me know if I can help in any way.

Statement on Diversity and Inclusion

As part of the [Smith School's commitment to diversity and inclusion](#), we recognize the importance of a diverse student body as necessary to a THRIVING environment. We are committed to fostering inclusive and equitable classroom environments. The Robert H. Smith School of Business strives to ensure all members of the Smith community feel welcomed, valued, and proud of every aspect of who they are. Through education, knowledge creation, advocacy, programming, and support, Maryland Smith commits to building an inclusive community that fosters a sense of belonging among all stakeholders.

I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

Technology Policy

Please refrain from using cellphones during class sessions. You **MUST** bring a laptop to class; we will use them to do frequent in-class activities.

Faculty Expectations

- I will respond to emails within 48 hours, M-F.
- Your email will get priority if you include the special tag [DMPA] in the subject header. I use this tag to make sure that I process class email first. If you forget to add the tag, and then remember after sending, just send it again using the tag.
- If you email me a question that I think will benefit everyone, I may ask you to post it on the Canvas discussion board before I respond.
- Feedback on assignments and grades will be returned in a timely manner, targeting an interval of 10 business days from submission.
- I will reserve questions sent to me in the 24 hours prior to the exam so I can share the answers with the class.

Student Expectations

Students who fail to adhere to these standards will see a negative impact on their grades and, in extreme cases, may be asked to leave.

- Be respectful of others and their contributions. Refrain from interrupting others, from holding side conversations during lectures, and from using disparaging tones.
- Join on time and stay through the end of the class period.
- It is your responsibility to check your email and ELMS at least once a day during the week (M-F), and I will expect you to be aware of any announcements within 24 hours of the time the message was sent. Contact me if you expect to be away from your computer for extended periods.

Attendance and Participation

- Students who miss a live session are responsible for learning what they miss from that session.
- Students must complete all readings and assignments in a timely manner in order to fully participate in class.

Absences and Late Policy

Late homework assignments require prior permission from the instructor and must be accompanied by a legitimate reason for not meeting a target deadline. Late assignments without permission will be penalized by 10% or a letter grade for each late 24-hour period, e.g., an assignment with a score of 15 will receive a score of 13.5 if it is 5 minutes late up to 24 hours late. Assignments that are more than 48 hours late will not be accepted.

The complete university policy on absences can be found at: <https://policies.umd.edu/assets/section-v/V-100G.pdf>.

For Major Scheduled Graded Events (MSGEs), only University approved absences (e.g., religious observance, unforeseen life circumstances out of your control (with documentation), or documented illnesses with appropriate documentation) will be accepted. Make-up exams are at my discretion.

Grade Appeal

Submit the graded assignment with a written statement explaining your request for re-grading. Justify the appeal by referring to specific course material and/or the syllabus. Appeals must be submitted within 7 days after a graded assignment has been returned.

If you appeal your grade, the entire assignment will be re-graded so the grade may increase or decrease. **Grade appeals will be considered at the end of the semester, only if the requested change in points would make a change in your final letter grade.**

Academic Integrity

The University's [Code of Academic Integrity](#) is designed to ensure that the principles of academic honesty and integrity are upheld. In accordance with this code, the Smith School does not tolerate academic dishonesty. Please ensure that you fully understand this code and its implications because all acts of academic dishonesty will be dealt with in accordance with the provisions of this code. All students are expected to adhere to this Code. It is your responsibility to read it and know what it says, so you can start your professional life on the right path. **As future professionals, your commitment to high ethical standards and honesty begins with your time at the Smith School.**

It is important to note that course assistance websites, such as CourseHero, are not permitted sources for Smith School courses [including this one](#). Material taken or copied from these sites can be deemed unauthorized material and a violation of academic integrity. These sites offer information that might not be accurate and that shortcut the learning process, particularly the critical thinking steps necessary for college-level assignments.

Additionally, it is understandable that students may use a variety of online or virtual forums for course-wide discussion (e.g., GroupME or WeChat). Collaboration in this way regarding concepts discussed in this course is permissible. However, collaboration on graded assignments is strictly prohibited unless otherwise stated. Examples of prohibited collaboration include: asking classmates for answers on quizzes or exams, asking for access codes to clicker polls, etc.

Finally, on each exam or assignment you must write out and sign the following pledge:

"I pledge on my honor that I have not given or received any unauthorized assistance on this exam/assignment."

Please visit the [Office of Undergraduate Studies' full list of campus-wide policies](#) and follow up with me if you have questions.

Course Evaluation

Please submit a course evaluation through CourseEvalUM in order to help faculty and administrators improve teaching and learning at Maryland. All information submitted to CourseEvalUM is confidential. Campus will notify you when CourseEvalUM is open for you to complete your evaluations for fall semester courses. Please go directly to the website (www.courseevalum.umd.edu) to complete your evaluations. By completing all of your evaluations each semester, you will have the privilege of accessing through Testudo, the evaluation reports for the thousands of courses for which 70% or more students submitted their evaluations.

For Prospective RAs, etc.

I receive a large number of requests to consider students and RAs. Please feel free to send me email, and I eventually will take at least a quick look through it. However, unfortunately, I do not have the time to respond to every email. If you have a potential research collaboration in mind, please see the details given on my website at <http://www.jessicamarieclark.com/for-students> for what I require to begin such a collaboration.

Regarding Letters of Recommendation, References, etc.

I am delighted to serve as a letter-writer for outstanding students who I have taught in class. "Outstanding" does not necessarily mean students who received an A. Make sure you have good attendance and participation in my class, treat me and other students with respect, and put forth a good effort on all assignments. Attending my office hours at least once is a good way to make sure I get to know you personally. Academic integrity and professionalism are of the utmost importance to me. I will not write a letter for someone who falls short on these! Please see the details given on my website at <http://www.jessicamarieclark.com/for-students> for more on what I require as a letter-writer.

Easter Egg

Your first assignment is to read and understand the syllabus by the second week of class. Please send a picture of the cartoon character you most identify with to jmclark@umd.edu from your UMD email with the subject header **[DMPA] 2022 Easter Egg**, in order to confirm that you did indeed read and understand the information given in this entire syllabus.

Copyright Notice

Course materials are copyrighted and may not be reproduced for anything other than personal use without written permission.

About the Instructor

I am an Assistant Professor of Information Systems. I have a PhD in Information Systems from the NYU Stern School of Business and a bachelor's degree in Math from Worcester Polytechnic Institute. I study best practices for using and interpreting machine learning in business analytics, focusing on applications such as advertising, social media, and crowdfunding. Data mining is my favorite subject to teach!