Washington University in St. Louis

Olin Business School

Course Syllabus - Fall 2022 Sections 21 to 33

COURSE NUMBER AND TITLE: DAT 500 S – Machine Learning Tools for Prediction of Business Outcomes

LAB MEETING TIMES (Central Time):

Section 21: T: 10:00 am – 11:20 am.
Section 22: T: 11:30 am – 12:50 pm.
Section 23: T: 1:00 pm – 2:20 pm.
Section 24: W: 10:00 am – 11:20 am.
Section 25: W: 11:30 am – 12:50 pm.
Section 26: W: 1:00 pm – 2:20 pm.
Section 27: Th: 10:00 am – 11:20 am.
Section 28: Th: 11:30 am – 12:50 pm.
Section 29: Th: 1:00 pm – 2:20 pm.
Section 30: Th: 2:30 pm – 3:50 pm.
Section 31: Th: 4:00 pm – 5:20 pm.
Section 32: Th: 5:30 pm – 6:50 pm.
Section 33: Th: 7:00 pm – 8:20 pm.

(You are expected to attend the lab in which you are enrolled in; you will be working with your group in every lab)

LOCATION: BH 210 N. **Remote option is available <u>ONLY</u> if you are cleared to stay remote by the school.**

OFFICE (Prof. Gerald Onwujekwe) Simon Hall 276 (In the office mostly on Thursdays)

OFFICE HOURS (Prof. Gerald Onwujekwe)

Tuesday 9:00 am – 10:00 am;

https://wustl.zoom.us/j/97418777189?pwd=aUdzMkdiWWVJUEorZ3ljalh6UzB2UT09

Meeting ID: 974 1877 7189

Passcode: 477327

By appointment during other times.

You are encouraged to ask conceptual questions to the Professor. It is encouraged that you ask R related (code de-bugging) questions to TAs.

CONTACT INFORMATION (Prof. Gerald Onwujekwe):

Phone- 314-935-6856

Email- geraldo@wustl.edu

OFFICE (Prof. Durai Sundaramoorthi) Simon Hall 265

OFFICE HOURS (Prof. Durai Sundaramoorthi)

Monday 1:30 pm – 2:30 pm;

https://wustl.zoom.us/j/93347390145

By appointment during other times.

You are encouraged to ask conceptual questions to the Professor. It is encouraged that you ask R related (code de-bugging) questions to TAs.

CONTACT INFORMATION (Prof. Durai Sundaramoorthi):

Phone- 314-935-3317

Email- sundaramoorthi@wustl.edu

TA Office Hours: 8:00 pm – 10:00 pm on M, T, W, Th, F, S, and Su

In addition to the instructor, TAs hold office hours for offering clarifications on the material, R programming, and homework. You are encouraged to ask conceptual questions to the Professor. It is encouraged that you ask R related (de-bugging) questions to TAs.

https://wustl.zoom.us/j/92978192178 Meeting ID: 92978192178

OLIN'S PILLARS OF EXCELLENCE:

OLIN STUDENTS WILL:

1. **Embody a values-based and data-driven ethos in their approach to all business situations**
2. Understand the global opportunities and challenges facing businesses
3. Engage with business issues through the application of experiential knowledge, in addition to the rigorous technical skills acquired in the classroom
4. Pursue world-changing initiatives with an entrepreneurial and innovative mindset and skillset

REQUIRED TEXT:

An Introduction to Statistical Learning with Application in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer. **(First Edition) (We are not using the second edition)**

Link to the book: https://www.statlearning.com/

Data Mining for Business Analytics: Concepts, Techniques, and Applications in R, Galit Shmueli, Peter C. Bruce, Inbal Yahav, Nitin R. Patel, Kenneth C. Lichtendahl WILEY Publishing.

SOFTWARE:

RStudio

COURSE OBJECTIVES:

We are living in the era of data abundance. Technological advancements in the last two decades in data gathering and storage have resulted in creation of Big Data. The challenge is to convert the massive amounts of data into useful information for decision making.

Upon completion of this course, you will acquire the fundamental knowledge in predictive analytics for analyzing data sets. You will be able to build predictive models useful for making real world business decisions. This course will raise your consciousness regarding the reality of the business data that you see. You will learn how to determine if what you are observing follows historical pattern, or if it could be explained by chance. This skill is extraordinarily useful in assessing customers' future behavior; for example, predictive analytics can help businesses to understand:

- Customer Segmentation
- Targeted Marketing
- Customers' reaction to price, promotion, display, etc.
- Consumers' basket
- Loans at higher risk for default
- Factors influencing the housing market
- Predictors of Salary
- Transactions that are fraudulent

You will become both an intelligent dispenser and consumer of predictive analytics. You'll learn what predictive analytics can and can't do ... how to spot the vulnerabilities of models that we create ... how to probe and intelligently challenge the "facts and arguments" disseminated by your competitors, your own company, and your elected governmental representatives.

SPECIFIC LEARNING OBJECTIVES:

(a) Understand the *language* of predictive analytics to communicate effectively with both experts and colleagues

(b) Understand the bias-variance tradeoff and impact of it on predictive models

(c) *Master* predictive analytics tools, applying them correctly to solve business problems. Among these tools are

- Classification Trees& Regression Trees
- Random Forest (Committee of Trees)
- Boosted Trees (This is a good competitor for Neural Network)
- Neural Network
- Logistic Regression
- Multi Adaptive Regression Splines
- Association Rules
- K-Nearest Neighbors
- Principal Component Analysis
- K-means Clustering
- Hierarchical Clustering

(d) *Develop Proficiency* in utilizing software to build models

(e) *Analyze* current business problems using the above tools

(f) *Evaluate* predictive models built by recognizing potential pitfalls

TENTATIVE COURSE SCHEDULE:

Lab	*Lab Date*	*Topic*	*Before Lab (Videos)*	*During Lab (Lab Topic, Assignments, etc.)*	*After Lab: (Assignments)*
	Aug 30, 31, & Sept 1	Discussion of Syllabus and creation of groups.			
1	Sept 6, 7, & 8	Module 1: Introduction to Machine Learning	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 2.4; Problem 1, Problem 2.	Group Assignment: Exercise 2.4; Problem 9	Individual Assignment: Exercise 2.4; Problem 10
2	Sept 13, 14, & 15	Module 2a: Model Accuracy Module 2b: Linear Regression	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 2.4; Problem 3; Problem 7. Exercise 3.7; Problem 3a and 3b; Problem 4.	Group Assignment: Exercise 3.7; Problem 9	Exercise 3.7; Problem 10
3	Sept 20, 21, & 22	Module 3a: Logistic Regression Module 3b: Linear Discriminant Analysis	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 4.7; Problem 1; Problem 4;	Group Assignment: Exercise 4.7; Problem 11 (leave out part e); Problem 12;	Individual Assignment: Exercise 4.7; Problem 10 (leave out part f)
4	Sept 27, 28, & 29	Module 4a: Quadratic Discriminant Analysis (QDA) Module 4b: Resampling Methods	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 4.7; Problem 5;	Group Assignment: Exercise 4.7; Problem 11 (part e); Exercise 5.4; Problem 5.	Individual Assignment: Exercise 4.7; Problem 10 (part f) Exercise 5.4; Problem 8
5	Oct. 4, 5, & 6	Module 5a: Bootstrap Module 5b: Subset Selection	Watch Concept and R Videos; Complete the online quiz of the module.	Group Assignment: Exercise 5.4; Problem 2 (part h); Problem 6;	Individual Assignment: Exercise 6.8; Problem 8 (parts a, b, c, d).

			Quiz: Exercise 5.4; Problem 2 (leave out h); Exercise 6.8; Problem 1.		
	Oct. 11, 12, & 13	No Lab	Preparation for Midterm Exam		
	Oct 14	This course will have a **Midterm Exam on October 14, 2022 (Friday) from 1 pm to 2:30 pm (Central) for all sections.** Every student enrolled is expected to appear for the exam in-person. However, students that obtained permission from the school to study remotely are expected to take the exams online at the same times listed above.			
	Oct. 18, 19, & 20	No Lab	Fall Break		
6	Oct. 25, 26, & 27	Module 6: Regularization	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 6.8; Problem 2; Problem 3.	Group Assignment: Exercise 6.8; Problem 9 (Parts a, b, c, and d).	Individual Assignment: Exercise 6.8; Problem 8 (Parts e and f)
7	Nov 1, 2, & 3	Module 7: Classification & Regression Trees	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 8.4; Problem 1; Problem 3; Problem 4.	Group Assignment: Exercise 8.4; Problem 9	Individual Assignment: Exercise 8.4; Problem 8 (Parts a, b, and c).
8	Nov 8, 9, & 10	Module 8a: Bagging and Random Forest Module 8b: Boosted trees	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 8.4: Problem 2; Problem 5;	Group Assignment: Exercise 8.4: Problem 12	Individual Assignment: Exercise 8.4: Problem 8 (parts d and e); Problem 10.
9	Nov 15, 16, & 17	Module 9: Support Vector Machines	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 9.7; Problem 3.	Group Assignment: Exercise 9.7; Problem 5.	Individual Assignment: Exercise 9.7; Problem 8.

	Nov 22, 23, & 24	No Lab	Thanksgiving Break		
10	Nov 29, 30, & Dec 1	Module 10: Clustering	Watch Concept and R Videos; Complete the online quiz of the module. Quiz: Exercise 10.7; Problem 2.	Group Assignment: Exercise 10.7; Problem 11.	Individual Assignment: Exercise 10.7; Problem 9
	Dec 6, 7, & 8	Review Class			
	Dec 12		This course will have a **Comprehensive Exam on Dec 12, 2022 (Monday) from 10 am to 11:30 am (Central) for all sections.** Every student enrolled is expected to appear for the exam in-person. However, students that obtained permission from the school to study remotely are expected to take the exams online at the same times listed above. Remote students are expected to have a high-speed connection and webcam on your computer. Graded exam 2 will not be returned to students, however, students are allowed to access and review their exam script in the presence of TAs on December 14. Please note that other students cannot access your exam.		

Attend the lab session in which you are enrolled. Before coming to the lab session, you are expected to watch all the concept and R videos posted for that module. You are also expected to complete the on-line quiz posted on Canvas for that module. **The deadlines for watching the videos and completing the associated quizzes on Canvas are at 11:59pm on Tuesdays. In addition to completing the Canvas quiz, you are expected to complete the quiz problems identified on the above schedule before 11:59 pm on Fridays after the lab.** The first 15 - 20 minutes of the lab time will be used for reviewing the content covered on the video and answering conceptual questions. The rest of the lab time will be used to work on the group assignment question posted on the above schedule. Fifteen minutes prior to the end of the lab session, one student from each group should give a quick one-minute report on their progress. They should also mention the difficulties faced by their group. During the last 10 minutes of the lab time, I will present some hints to arrive at the solution. **The group assignment report is due at 11:59 pm on Saturdays after the lab sessions.** In every module, students are assigned to work on an assignment individually. **The individual assignments are due at 11:59 pm on Sundays after lab sessions.** The individual assignment questions are also identified on the above schedule. If a group completes their group assignment during the lab, students should utilize the rest of the lab session to work on the individual assignment for that module. While working on the individual assignments, students are allowed and encouraged to discuss the problems they face, potential solutions to problems, and coding strategy with other fellow students. **However, students are not allowed to copy code or report from other students. After sharing ideas, everyone is expected to write their own code and produce and submit reports on their own.** The report should identify all the collaborators and the nature of the collaboration. **Copying reports and code from other students, solution manuals, and online resources will be considered as a violation of the honor code. Students are prohibited to utilize solutions posted from the prior years. Students are prohibited from obtaining a copy of the solution from other students or TAs. You are welcome to utilize the solution hints posted by the instructor on Canvas or presented in the lab.**

GRADING:

- Online Quizzes (Multiple Choice; Due on Tuesdays) 10%
- Quizzes (Textbook Questions; Due on Fridays after lab) 10%
- Group Assignments+ (Due on Saturdays after lab) 15%
- Individual Assignments (Due on Sundays after lab) 15%
- Attendance & Participation* 10%
- Exam 1 (1pm to 2:30pm (Central) on October 14, 2022 (Friday) 20%
- Exam 2 (10 am to 11:30 am (Central) on Dec 12, 2022 (Monday) 20%

*** Remote Students are expected to have a high-speed internet connection and webcam during the lab sessions and exams.**

+ Each group should have five or six members. Contributing adequately to the group assignment is required to obtain a passing course grade. Students should inform me if some members are not contributing adequately to the group assignments.

Students taking the course for pass/fail are expected to participate in every group assignment, individual assignment, quiz, and exam in order to pass the course. In addition, scoring 75% or above is required for a pass. Scoring 90 % or above is required for a high pass.

To avoid missing participation grade, students participating in university organized events are encouraged to avoid conflicts with the lab schedule. During every lab session, students are graded for their attendance and participation. Typically, make-up exams, quizzes, and assignments are not given.

Late submissions are penalized as following: 10% penalty will be applied for late submissions made within 24 hours of the due date; 25% penalty will be applied for late submission made after 24 hours but within 48 hours of the due date; 50% penalty for late submissions made after 48 hours but within a week from the due date; late submissions are not considered after a week.

How to Succeed in this Course:

(a) Attend each lab session and be actively engaged. If you are not engaged, talk to the instructor during office hours to find ways to be excited about this course!
(b) Open yourself to learning the value of mathematical symbolism, predictive modeling techniques, analytic arguments, and tight, self-consistent answers.
(c) Devote considerable time to the assignments.
(d) Practice exercise problems for each topic.
(e) To become familiar with R about five hours of additional effort beyond class time per week is needed.
(f) Finally, if something doesn't make sense during the class, ask a question. Don't worry about what others may think, as they are probably unclear as well.

ATTENDANCE: I am committed to make the lab time worthwhile for you. Students are expected to attend every lab, arrive to class on time, remain engaged in class, and be prepared to participate (e.g., bring textbook, laptop, notes, calculator, and be ready for discussions). If you are attending the class on zoom, you are expected to have high-speed internet and webcam for every lab and exams. Students are expected to take turns to present their status-report in the class. Students will be held responsible for any content covered in the class and the video. If you know you are going to miss a lab for unavoidable reason, send an email to the instructor as soon as possible and save instructor's response for your records. If you happen to miss a class for unforeseen/unexpected but acceptable reasons, you are expected to email and inform the instructor as soon as possible. **During every lab, students will be graded for attendance and participation. If you are a remote student joining on zoom, you are expected to have your camera turned on during the entire lab session/exam.**

MAKE-UP POLICY: Typically, the deadlines are strict and no make-up opportunities are given beyond the late submission policy mentioned before. A submission can be re-scheduled provided a notice is given and accepted well before the deadline. Students who have excusable absences should contact the instructor to request special arrangements as soon as possible.

CLASSROOM CONDUCT: A positive learning environment depends upon mature behavior. Among other things, this means arriving to class on time, listening quietly and respectfully when someone else is speaking, and always using socially acceptable business language. Appropriate steps will be taken to ensure this positive learning environment.
Usage of tablets and cellphones (e-mailing, texting, chatting, browsing internet, browsing social media content, working on other assignments, playing computer games, etc.) during the class is

prohibited for other purposes apart from how the instructor has instructed to use these devices. Taking rest room breaks as a habit during the lab is discouraged.

THINGS THAT ARE NOT EXPECTED TO BE USED IN CLASS: Cell phones, iPods, etc. A cell phone cannot be used to replace a calculator. If you need to tape the lab, you need to obtain instructor's prior approval.

Confidential Resources for Instances of Sexual Assault, Sex Discrimination, Sexual Harassment, Dating Violence, Domestic Violence, or Stalking:

If a student needs to explore options for medical care, protections, or reporting, there are free, confidential support resources and professional counseling services are available through the Relationship and Sexual Violence Prevention (RSVP) Center in Seigle Hall, Suite 435, rsvpcenter@wustl.edu, 314-935-3445. For after-hours emergency response services, call 314-935-6666 or 314-935-5555 and ask to speak with an RSVP Counselor on call.

Academic Accommodations:

Reasonable Accommodations for Disabled Students - Washington University in St. Louis supports the rights of enrolled students to a full and equal educational opportunity and, in compliance with federal, state, and local requirements, is committed to reasonable accommodations for individuals with documented disabilities. Disabled students for whom accommodations may be necessary must be registered with, and provide their instructors official notification through, WUSTL's **Disability Resources** (https://students.wustl.edu/disability- resources/). Once established, responsibility for disability-related accommodations and access is shared by DR, faculty, and the student. Please contact Disability Resources at 314.935.5970 or disabilityresources@wustl.edu.

Sexual Assault Resources - The University is committed to offering reasonable academic accommodations (e.g., a no-contact order, course changes) to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If you need to request such accommodations, please contact RSVP (information above) to schedule an appointment with an RSVP confidential and licensed counselor. Although information shared with counselors is confidential, requests for accommodations will be coordinated with the appropriate University administrators and faculty. See: RSVP Center

Mental Health:

Mental Health Services' professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect a student's academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety, depression, and thoughts of suicide. See: https://students.wustl.edu/mental- health-services/ Additionally, see the mental health services offered through the RSVP Center listed above.

WashU Cares:

WashU Cares, within the Health and Wellness Unit, provides resources to all students on the Danforth Campus who may be having a hard time. WashU Cares is committed to helping create a culture of caring. Through proactive, collaborative, and systemic approaches, WashU Cares works with students to identify interventions, resources, and supports that allow them to be successful. If there is a concern about the physical or mental well-being of a student, please file a report on the WashU Cares website. See: https://washucares.wustl.edu/.

Center for Diversity and Inclusion (CDI):

The Center for Diversity and Inclusion (CDI) supports and advocates for undergraduate, graduate, and professional school students from underrepresented and/or marginalized populations, collaborates with campus and community partners, and promotes dialogue and social change to cultivate and foster a supportive campus climate for students of all backgrounds, cultures, and identities. See: https://diversityinclusion.wustl.edu/.

ACADEMIC INTEGRITY: I take matters of Academic Integrity seriously and expect that you will also. The Olin Business School is a community of individuals with diverse backgrounds and interests who share certain fundamental goals. Primary among these goals is the creation and maintenance of an atmosphere conducive to learning and personal growth for everyone in the community.

Becoming a member of the Olin community is a privilege which brings certain responsibilities and expectations. The success of Olin in attaining its goals and in maintaining its reputation of academic excellence depends on the willingness of its members, both collectively and individually, to meet their responsibilities. All individuals associated with Olin should conduct themselves with the utmost integrity in all aspects of their life, both on and off campus.

Please refer to the *Olin Code of Conduct handbook* for specific responsibilities, guidelines and procedures regarding academic integrity

I encourage you to ask me if you have any questions about any academic integrity issue that you may have, e.g., what's allowed and what's not. It is important to note that these are different for different courses.

(a) All individual assignments should reflect a significant effort of yours even though you are allowed to collaborate.
(b) Signing attendance for others or asking others to sign attendance for you is prohibited.
(c) Passing class notes, class handouts or assignments to students who have yet to take the course, who attend a different section or receiving material from other students is strictly prohibited.
(d) Similarly, having access to an archive of past course material is prohibited.
(e) Attending a different section of the class without instructor's prior approval is prohibited.
(f) Each and every member of the group is expected to contribute to the group assignment in an equitable manner. All group members are responsible for the academic integrity of the assignment.
(g) Plagiarizing (the misrepresentation of work done by others as being one's own work) is a violation of the Code of Conduct. Remember to cite all sources of information and ideas to prevent problems. It is highly discouraged to borrow sentences or phrases from other sources. If you cannot avoid borrowing such quotes from other sources, make sure to enclose every quotation within a quotation marks and acknowledge its source.
(h) If you use figures or tables from other publications, you need to cite the source right below those figures and tables.
(i) Submitting similar assignments in more than one class requires approval of all faculty members involved.
(j) Cheat sheets are not allowed for the exam.

You are responsible to notify me (anonymously if you wish), if you suspect that improper behavior has occurred. I will immediately initiate an investigation and if the charges have merit, I will bring the matter before the Disciplinary Committee.

As in the past, any proof of an Honor Code violation will result in swift academic disciplinary action.

COVID-19 Health and Safety Protocols:

While on campus, it is imperative that students follow all public health guidelines established to reduce the risk of COVID-19 transmission within our community. The full set of University protocols can be found at https://covid19.wustl.edu/health-safety/. This includes:

- Completing a self-screening using the WashU COVID-19 Screening app every day before coming to campus or leaving your residence hall room. If you do not receive a green check and pass the screening, you are not permitted to come to campus or leave your residence hall room. You must contact the COVID Call Center (314-362-5056) or the Habif Health and Wellness Center (314 935-6666) immediately. Note: In addition to the symptoms listed in the screening tool, everyone also should pay attention to symptoms that are new or different for you, including things like headache and congestion, particularly in combination with diarrhea. These can also be signs of COVID-19. Call the COVID Call Center or Habif to report these symptoms.
- Complying with universal masking. All individuals on campus must wear disposable masks or cloth face coverings while occupying indoor public settings, including: multi-person offices, hallways, stairwells, elevators, meeting rooms, classrooms and restrooms. Masks are encouraged but not required for outdoor activities, particularly at large events or in crowded settings.
- Maintaining physical distancing as needed. While distancing requirements have been removed for vaccinated students, those who are not fully vaccinated are strongly encouraged, for their own health, to maintain a distance of 6 ft from others in the classroom. If you are not able to be vaccinated or have conditions that may put you at increased risk of failed immunity and classroom activities would bring

you in frequent proximity to other students, contact your instructor to discuss alternatives.

- Practicing healthy personal hygiene, including frequent handwashing with soap and warm water for at least 20 seconds and/or using hand sanitizer with at least 60% alcohol.

What should I do if my instructor has been disconnected from Zoom?

- If we are in a hybrid or in-person format, the instructor (or the Classroom Engagement Moderator) will contact AV support to get the classroom logged back in as quickly as possible. Stay on the Zoom link and monitor your email for additional instructions.
- If we are in a fully virtual format, students should remain on Zoom for at least 10 minutes. Students may then exit the session, but they should ensure they have access to their email for an additional 15 minutes to allow the instructor the opportunity to provide updated communication and direction on how the class will continue. If the instructor does not return or provide updated communication after 15 minutes, the remainder of the class will be canceled then later completed asynchronously via a recording that will be posted on Canvas.

What should I do if I have been disconnected from Zoom?

- Students who have been unexpectedly disconnected from Zoom should work to reconnect utilizing the technology available to them at the time. If the student is unable to reconnect, the student should notify the instructor via their preferred method of communication and then use the class recording to make up for the lost material.