



ITAO 40250: Unstructured Data Analytics

Section 1: TR 9:30 AM - 10:45 AM

Section 2: TR 11:00 AM - 12:15 PM

Room: MCOB L003

Instructor

John Lalor (john.lalor@nd.edu)

Office: 338B Mendoza

Office Hours

Thursdays 12:30pm-2:00pm

MCOB 338B

Or by appointment

Teaching Assistants

Badar Cha (jcha2@nd.edu)

Grace Courtney (gcourtne@nd.edu)

Office hours

Mendoza basement

Wednesday 5:00-7:00pm (Grace, **L050**)

Thursday 5:00-7:00pm (Badar, **L050**)

Course Feedback Form

If at any point during the mod you would like to submit *anonymous* feedback regarding the course, you can do so at the following link: <https://forms.gle/AKTckonCyN9p5qcM7>.

Mission Statements

University of Notre Dame “The University of Notre Dame is a Catholic academic community of higher learning, animated from its origins by the Congregation of Holy Cross. The University is dedicated to the pursuit and sharing of truth for its own sake. As a Catholic university, one of its distinctive goals is to provide a forum where, through free inquiry and open discussion, the various lines of Catholic thought may intersect with all the forms of knowledge found in the arts, sciences, professions, and every other area of human scholarship and creativity.

The intellectual interchange essential to a university requires, and is enriched by, the presence and voices of diverse scholars and students. The Catholic identity of the University depends upon, and is nurtured by, the continuing presence of a predominant number of Catholic intellectuals. This ideal has been consistently maintained by the University leadership throughout its history. What the University asks of all its scholars and students, however, is not a particular creedal affiliation, but a respect for the objectives of Notre Dame and a willingness to enter into the conversation that gives it life and character. Therefore, the University insists upon academic freedom that makes open discussion and inquiry possible.”(source)

Mendoza College of Business “To build a premier Catholic business school that fosters academic excellence, professional effectiveness and personal accountability in a context that strives to be faithful to the ideals of community, human development and individual integrity.” (source)

This course To teach students how to manage unstructured data and use unstructured data for analytics, with a specific emphasis on business applications and ethical considerations.

Course Policies

Academic Honor Code Entering Notre Dame you were required to study the on-line edition of the Academic Code of Honor, to pass a quiz on it, and to sign a pledge to abide by it. The full Code and a Student Guide to the Academic code of Honor are available at: <http://honorcode.nd.edu>. From that page: “All members of the University community are expected to embrace and adhere to the following pledge:

As a member of the Notre Dame community, I acknowledge that it is my responsibility to learn and abide by principles of intellectual honesty and academic integrity, and therefore I will not participate in or tolerate academic dishonesty.”

The pledge to uphold the Academic Code of Honor includes an understanding that a student’s submitted work, graded or ungraded – examinations, draft copies, papers, homework assignments, extra credit work, etc. - must be his or her own.

On the Use of Generative AI Models The below comes from Notre Dame’s Statement Regarding Generative AI:*

*When students use generative AI (such as ChatGPT) to replace the rigorous demands of and personal engagement with their coursework, it runs counter to the educational mission of the University and undermines the heart of education itself. Artificial intelligence, large language models, and technologies hold promise for deploying knowledge in service to others and accelerating the discovery of new knowledge. However, such technology poses new challenges to pedagogy and to integrity. Within the context of the teaching mission of the University, and consistent with the Academic Code of Honor, **the authority to define the appropriate use, study, and deployment of these technologies rests with the faculty.***

The policy of this class with regards to Generative AI is that of **full disclosure**. If you use a Generative AI tool, you must disclose it. This includes the model used, prompts used to generate outputs, etc. You as the student are responsible for what you turn in as coursework.

Academic Accommodations It is the policy and practice of The University of Notre Dame to provide reasonable accommodations for students with properly documented disabilities. Students who have questions about Sara Bea Disability Services or who have, or think they may have, a disability are invited to contact Sara Bea Disability Services for a confidential discussion in the Sara Bea Center for Students with Disabilities or by phone at 574-631-7157. Because the University’s Academic Accommodations Processes generally require students to request accommodations well in advance of the dates when they are needed, students who believe they may need an accommodation for this course are encouraged to contact Sara Bea Disability Services at their earliest opportunity. Additional information about Sara Bea Disability Services and the process for requesting accommodations can be found at sarabeadisabilityservices.nd.edu.

Support for Student Mental Health Diminished mental health can interfere with optimal academic performance. The source of symptoms might be related to your course work; if so, please speak with me. However, problems with other parts of your life can also contribute to decreased academic performance. The University Counseling Center (UCC) provides cost-free and

*<https://honorcode.nd.edu/official-statement-regarding-generative-ai-may-2023/>

confidential mental health services to help you manage personal challenges that threaten your emotional or academic well-being.

Remember, getting help is a smart and courageous thing to do — for yourself and for those who care about you. For more resources please see ucc.nd.edu The UCC is located on the third floor of Saint Liam Hall

Phone: 574-631-7336.

Hours: Monday-Friday 8:30am – 5:00pm.

Urgent Crisis Line 24/7

Smart Use of Technology This course relies heavily on access to computers, specific software, and the Internet. At some point during the semester **you will have a problem with technology:** your laptop will crash, a file will become corrupted, a server will go down, or something else will occur. These are facts of life, not emergencies. Technology problems will not normally be accepted as excuses for unfinished work. Count on “stuff” happening and protect yourself by doing the following:

- Plan ahead – start early, particularly if scarce resources are required
- Save work often – at least every ten minutes
- Make regular backups of files in a different location from the originals
- Save drafts of work at multiple stages
- When editing an image, set aside the original and work with a copy
- Practice safe computing when surfing the web and checking email
- On your personal computer, install and use software to control viruses and malware

Course Description

Approximately 80% of the world’s data is unstructured, that is data that is not stored in a relational database. It is growing at fifteen times the rate of databases. Examples of unstructured data include e-mails, financial filings, customer feedback, blogs, online reviews, instant messages, tweets, pictures, and videos. Extraction of information from unstructured data is increasingly important for companies and other organizations but is still a nascent practice area within them. Analytic techniques are increasingly important for understanding what can be learned from unstructured data sets and demand is strong for unstructured data analytical skills. This course introduces the student to the basic concepts of unstructured data and the basic analytic techniques used to extract useful information from this data. The emphasis will be on determining what data to collect, how to collect it, how to process it and finally creating analytics from the data to solve basic business problems. Sources of text data will include emails, tweets and various documents. The analytics techniques will focus on solving problems using word counts, document term matrices, and classification techniques. The learning outcomes from this class will provide students with the basic skills sought in the marketplace.

Student Learning Goals

Upon successful completion of this course, you will be able to:

1. Describe the algorithms used in the analysis of unstructured data and compare them based on appropriateness for various business problems
2. Identify sources of unstructured data, assess the value of the data, and design a plan to go from raw data to useful analytics and visualizations
3. Implement a pipeline in Python to import, clean, perform analytics, and visualize a given data set
4. Debate the ethical implications of analytics tools in business

Integrating Virtues Initiative

This course is part of Notre Dame's Virtues & Vocations initiative, which promotes the integration of virtue ethics and moral purpose across academic disciplines. Concepts in this class are at the foundation of modern artificial intelligence (AI) and large language models (LLMs). Therefore, we will cultivate the moral virtues of *humility* and *hope*, the intellectual virtue of *curiosity*, and the civic virtue of *deliberation*. We will approach the course topics, and the rapidly changing landscape of artificial intelligence, with a humble and hopeful perspective. Curiosity will guide our investigation of the concepts and techniques so that we can deeply understand how and why AI systems generate outputs the way that they do. Finally, deliberation will help us put what we learn into a broader perspective, forcing us to “zoom out” from the coding exercises to consider larger implications. We will incorporate this virtue cultivation through assigned readings, in-class discussions, and our final project.

Prerequisites

I expect that you have taken ITAO 30220 (Predictive Analytics), ITAO 30230 (Data Management), and have taken ITAO 20210 (Coding Fundamentals) OR are currently taking ITAO 40230 (Data Analysis with Python). There will be a lot of Python in this course, and I am assuming a basic familiarity on your part. If you have not taken Python, or if you find yourself struggling with the assignments, please see me.

Course Materials

Course materials will be posted on **Canvas** as needed. The following textbooks are available for free online, and are useful resources for brushing up on your Python. Some readings will be taken from these books:

- Automate the Boring Stuff with Python: <https://automatetheboringstuff.com/>
- Python Data Science Handbook: <https://jakevdp.github.io/PythonDataScienceHandbook/>
- Natural Language Toolkit (NLTK) Book: <https://www.nltk.org/book/>

Course Requirements

Final Exam The final exam will consist of multiple choice and free-response questions to assess understanding of concepts covered during the course. The final exam is **open notes**. Please refer to the schedule below for dates and times of the exam.

Group Project There will be a group project that will consist of (1) a short report (3-4 pages) and (2) a presentation (7-10 minutes). The goal of the group project is to go from an unstructured data set *in the wild* to a set of actionable insights using the techniques described in class. Final projects will be graded based on application of the concepts covered in class, interestingness of the problem/data source, and final report/presentation quality. I expect the report to include sections on data identification and cleaning, identifying and analyzing structure in the data, description of the analytical problem, and results. More details on the project requirements will be posted on Canvas.

Groups of 3-4 students will be randomly assigned. There will be peer evaluation as part of the group project, assigned in the middle of the course, to check progress on the project and to account for individual effort in the project.

Assignments There will be four (4) at-home assignments. Each assignment is worth 10% of your final grade. The assignments will be programming assignments where you will be asked to implement concepts discussed in class in Python. For each assignment you will be required to turn in a PDF rendering of a Jupyter notebook. I will provide the initial Jupyter notebook with assignment requirements and questions prior to the assignment due date.

All assignments will be turned in on **Canvas** and will be due by **11:59pm of the due date**. **Canvas** timestamps homework assignments, and also allows me to lock submission after a certain period. Students will be allowed two (2) late days that can be used for any assignment. Once you have used your late days, any future late submissions will receive a grade of 0. No assignment submissions will be accepted via email.

Note: The free days for an assignment are all or nothing. If you submit Assignment 1 at 12:01am on the day after the due date, you will be marked as having used one late day. If you then submit Assignment 2 at 11:59pm the day after the due date, you will be marked as having used your second and final late day. The rest of the assignments must be submitted on time to receive credit.

Each student must complete and submit his or her own assignment. You may discuss strategies for the assignments with other students but group work is not permitted.

Attendance and Participation Attendance is expected, as per the policy noted above. "Prep materials" (readings and/or notes) will be posted to **Canvas** before each class, and it is my expectation that you will have read all prep materials before each class. Section 3.1 of the Faculty Handbook (<http://facultyhandbook.nd.edu/governance/>) outlines the University's policy on excused absences. The students are expected to attend classes and are responsible for obtaining information about missed classes from other students, such as changes to the due dates and contents of exams, assignments and projects. They are also expected to actively participate in all class discussions. Student participation will be a key component of the class. Students are expected to ask questions and participate in discussions about the course material.

Grading Your final grade will be comprised of the following elements, listed in order of the percent contribution to the final grade:

1. Assignments: 40%
2. Final exam: 25%
3. Final project: 25%
4. Attendance and Participation: 10%

Letter grades will be assigned based on the scores earned on the items listed above, and will follow the following scale:

Range	Grade	Range	Grade
93 - 100	A	77-79.9	C+
90 - 92.9	A-	73 - 76.9	C
87 - 89.9	B+	70 - 72.9	C-
83 - 86.9	B	60 - 69.9	D
80 - 82.9	B-	0 - 59.9	F

Please note that this scale may be adjusted to ensure that the mean (average) final grade for this course is within the range 3.0 - 3.6 as per the ITAO department grading policy.

Course Schedule

Note: this schedule is subject to change. Any changes will be communicated to students in a timely manner via Canvas.

Week	Date	Topic	Activity/Assignment
Section 1: Getting the data, seeing the data			
1	10/16/2025	Course Overview, Data Analysis in Python, Integrating Virtue Introduction	A0 assigned
Fall break: 10/20-10/24			
2	10/28/2025	Data Collection and Preparation, Text Analysis	A0 due 10/26, 11:59 PM
Section 2: Unsupervised Analytics			
	10/30/2025	Visualization, Dimensionality Reduction	A1 assigned
	11/04/2025	Clustering	A1 due 09/03, 11:59pm
3	11/06/2025	Topic Modeling, Word Embeddings	A2 assigned
	11/11/2025	Project Progress Update	A2 due 09/10, 11:59pm
Section 3: Supervised Analytics			
4	11/13/2025	Text Classification, Sentiment Analysis	A3 assigned
	11/18/2025	Text Classification Part 2 Introduction to Neural Networks	A3 due 09/17, 11:59pm
5	11/20/2025	Neural Networks Part 2	A4 assigned
	11/25/2025	Optical Character Recognition (OCR) Image Classification	A4 due 09/24, 11:59pm
Section 4: Looking Forward			
6	11/27/2025	Ethics, Integrating Virtue Revisited	
	12/02/2025	Emerging Topics	
7	12/04/2025	Final Presentations	
	12/09/2025	Final Presentations	

Final Exam: Wednesday, December 17th 8:00am - 10:00am, **DeBartolo 141**