Course Syllabus: Data Mining and Text Analytics in Business

Dillard College of Business Administration
MIS 5613 Section Y20
Spring Semester 2026

Contact Information

Instructor: Jiaxi Luo
Office: Dillard Building 216
Office hours: Mondays & Wednesdays 3:30 – 5:30p Tuesdays 10:00a-11:00a
Office phone: (940) 397-3268
E-mail: jiaxi.luo@msutexas.edu

Course Description

The course will deliver concepts and skills of the various techniques and methodologies used in data mining and text analytics, including supervised and unsupervised learning: decision trees, clustering, regression, support vector machines, and neural networks; text mining techniques such as sentiment identification, topic classification, text summarization, and text generation under realistic time constraints.

Students will learn how to apply these techniques to solve real-world business problems such as customer segmentation, sentiment analysis, and predictive modeling. The course will also provide hands-on experience using popular data mining tools: Python to extract insights from large datasets.

In addition, students will be introduced to the latest trends and advancements in text analytics, including machine learning-based approaches for text classification, topic modeling, and named entity recognition. By the end of the course, students will have a comprehensive understanding of data mining and text analytics and be able to apply their knowledge to real-world business scenarios.

Course Specific Learning Goals: After completing this course, students should be able to:

- Understand the fundamental concepts and techniques of data mining and text analytics, including text pre-processing, feature extraction, and classification.
- Learn how to use Python libraries such as pandas, NumPy, and scikit-learn for data manipulation and analysis.
- Learn how to use Python libraries such as NLTK and spaCy for natural language processing tasks, such as tokenization, stemming, and lemmatization.
- Understand how to apply text analytics and data mining techniques to real-world business problems, such as sentiment analysis, topic modeling, and predictive modeling.
- Learn how to visualize and communicate the results of data mining and text analytics using Python libraries such as matplotlib and seaborn.
- Get hands-on experience with a variety of real-world text datasets and business case studies.
- Understand the ethical and legal considerations of data mining and text analytics, and the importance of responsible data practices.

Textbook & Instructional Materials

Suggested Reading Lists:

Machine Learning with Python Cookbook; ISBN-13: 978-1491989388 by Chris Albon
Applied Text Analysis with Python; ISBN-13: 978-1491963043 by Benjamin Bengfort, Rebecca Bilbro, and Tony Ojeda
Machine Learning Techniques for Text: Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation; ISBN-13: 978-1803242385 by Nikos Tsourakis
Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2; ISBN-13: 978-1789955750 by Sebastian Raschka, Vahid Mirjalili
Python Machine Learning By Example: Build intelligent systems using Python, TensorFlow 2, PyTorch, and scikit-learn; ISBN-13: 978-1800209718 by Yuxi Liu

A PC/laptop/tablet with webcam capability, be able to run PyCharm Professional (Chromebooks won't work due to insufficient computing power)
PyCharm Professional (Free for student account)
Additional readings are posted to D2L.

Student Handbook

Refer to: Student Handbook

Academic Misconduct Policy & Procedures

Academic Dishonesty: Cheating, collusion, and plagiarism (the act of using source material of other persons, either published or unpublished, without following the accepted techniques of crediting, or the submission for credit of work not the individual's to whom credit is given). Additional guidelines on procedures in these matters may be found in the Office of Student Conduct.
Office of Student Conduct

Moffett Library

Moffett Library provides resources and services to support student's studies and assignments, including books, peer-reviewed journals, databases, and multimedia materials accessible both on campus and remotely. The library offers media equipment checkout, reservable study rooms, and research assistance from librarians to help students effectively find, evaluate, and use information. Get started on this Moffett Library webpage to explore these resources and learn how to best utilize the library.

Grading

Points will be allocated using the following scheme. Grades will be based on the recorded points only. Personal reasons (e.g., need a specific grade to graduate, to keep financial aid, to keep a straight A record, etc.) are not considered in the grade calculation.

Table 1: Points allocated to each assignment

Assignments	Points
Exam I	66
Exam II	67
Exam III	67
Project	150
Quizzes	100
Homework	100
Class Participation	50
Total	600

Table 2: Total points for final grade.

Grade	Points
A	540-600
B	480-539
C	420-479
D	360-419
F	<360

Homework

Homework consists of Python programming assignments and counts **100** points toward the course grade. From the Data Mining portion onward, most assignments include a subjective interpretation component.
How and when to turn it in:

- **Format**: Submit standard .**py** files—one .**py** per question—and keep the .**py** files as pure code. For the interpretation part, you may either (a) write it as comments inside the corresponding .**py**, or (b) place all interpretations in a single Word document (.**docx**) and clearly reference the related .**py** filenames.
- **Where**: Upload to the D2L Dropbox for the corresponding assignment.
- **When** (due time): Midnight (11:59 PM) on the due date; D2L timestamp is official.
- **Late work**: Accepted up to one week (7 calendar days) after the deadline with a grade penalty (see assignment page for details).
- **Verify the file opens correctly.**

Quizzes

Up to **100** points will be assigned to quizzes. Quizzes are delivered in D2L Quizzes on a regular basis to reinforce each section's skills; most items are auto-graded (occasional short answers may require manual review). Each quiz has a posted time limit and an availability window; complete it by midnight (11:59 PM) on the due date—late submissions after the window closes are not accepted. Quizzes are individual work and subject to the course academic-integrity policy; unless otherwise noted, you may use your notes and the internet for reference, but collaboration or answer-sharing is prohibited. Scores and answer keys may be released after the window closes as indicated on D2L. Any exceptions (e.g., make-ups, accessibility accommodations) must follow university policy and be arranged in advance; see D2L for exact timing, attempt limits, and any drop/replace-lowest rules announced for the term.

Exams

There will be three online exams. On each posted exam date, you may begin any time within the availability window; once you start, you have **5 hours** to complete the exam, plus a 5-minute grace period to save and submit. Each exam contains 20 **multiple-choice** questions in D2L Quizzes and 5 **coding questions** completed outside the quiz. This is an **open-book**, **open-notes**, **open-internet** exam; however, work is strictly individual. Academic-integrity violations will be referred to the Office of Student Conduct and may result in a failing grade for the course and additional disciplinary actions, in accordance with university policy.
For multiple-choice, respond directly in D2L Quizzes. For coding, keep the quiz open and navigate to D2L → Assignments → Exam Dropbox Folder to download any required files. Complete your Python in any IDE and submit one **.py** file per question; place any short answer/interpretation either as comments inside the

corresponding **.py** or in a single **PDF/Word/TXT** that clearly references the **.py** filenames. You may use AI for coding, but all subjective responses must be written by you and grounded in your results; if AI is used, add 2–4 sentences per question explaining what you verified/changed and why. Do not change required file/column names and use only the model specified in each question.
If the instructor has questions or concerns about the originality of a student's responses, the student may be required to meet with the instructor during office hours to provide a verbal explanation of their answers. The outcome of this meeting may be used to determine or adjust the exam's final score.

Projects Required

Teams of 3–4 students will complete a mini-project in data mining and text analytics. Each team must obtain its own dataset from a public repository or collect one via web crawling (crawling earns extra credit if done ethically and in accordance with site policies). The project should demonstrate a full workflow—data preprocessing; classification, regression, or clustering; model selection; and evaluation—with a strong emphasis on text analytics (e.g., sentiment analysis, topic classification, summarization). Your Project Report should cover: a concise dataset summary (source, topic, size, attributes), preprocessing steps with before/after examples, EDA with descriptive statistics and core plots, train/test split strategy, the rationale for your chosen task and model(s), feature selection (if used), the chosen text analysis approach and justification, evaluation methods and interpretation, optional model visualizations, and actionable findings and recommendations. Examples of data sources include Kaggle, UCI ML Repository, Tableau Public sample sets, Wikidata, Google Dataset Search, and Makeover Monday.
Submission: Upload to the D2L Dropbox by the posted deadline:
(1) a 10–15 minute recorded presentation highlighting insights and proposed solutions; (2) the Project Report (PDF/Word); and (3) all project code and data (submit .py files and any required .csv/artifacts; if you crawled data, include the crawl scripts and a brief note on how you complied with site policies). See D2L for exact formatting and the due-time policy.

Class Participation

Class Participation is based primarily on students' contributions to required **D2L online discussions**, including the quality, relevance, and timeliness of discussion posts and responses. Expectations and grading criteria for discussions will be explained in class and posted on D2L.

Late Work

Late homework, less than one week after the deadline, may be accepted; certain points will be deducted.

Make Up Work/Tests

Students with excused absences may make up missed examinations and in-class activities, but supporting documents are required. Arrangements must be made in advance if possible. In all cases, the instructor must be contacted no later than the day of the scheduled exam, or no make-up will be allowed. At the instructor's discretion, a deduction may be assessed for a late exam.
Excused absences include active military/police/firefighter assignment, jury duty, university-authorized absences (for example, athletic events or study-abroad programs), and medical emergency for yourself or your immediate family member. For more information about university-authorized absences, please refer to the Midwestern State University Undergraduate Catalog:
https://catalog.msutexas.edu

Important Dates

- Last day for term schedule changes: 01/20/2026 – 01/23/2026.
- Deadline to file for graduation: 02/16/2026
- Last Day to drop with a grade of "W:"04/29/2026
- Check the date on the Academic Calendar
- Refer to: Drops, Withdrawals & Void

Attendance

Students are expected to attend all meetings of the classes in which they are enrolled. Although students are generally graded on intellectual effort and performance rather than attendance, absences may lower a student's grade when class attendance and participation are deemed essential by the faculty member. In classes where attendance is considered part of the grade, the instructor should inform students of the specifics in writing at the beginning of the semester, in a syllabus or a separate attendance policy statement. An instructor with an attendance policy must keep daily records. The instructor must give the student a verbal or written warning before dropping the student from the class. Instructor's records will stand as evidence of absences. A student with excessive absences may be dropped from a course by the instructor. Any individual faculty member or college may establish an attendance policy, provided it is in accordance with the General University Policies. **Students who accumulate 5 or more unexcused absences in class sessions will lose all Participation points.**

Desire-to-Learn (D2L)

Extensive use of the MSU D2L program is a part of this course. Each student is expected to be familiar with this program, as it serves as the primary source of communication for assignments, examination materials, and general course information. You can log into D2L through the MSU Homepage. If you experience difficulties, please contact the program's technicians or your instructor.

Online Computer Requirements

Taking an online class requires you to have access to a computer (with Internet access) to complete and upload your assignments. It is your responsibility to have (or have access to) a working computer in this class. ***Assignments and tests are due by the due date, and personal computer technical difficulties will not be considered a reason for the instructor to allow students extra time to submit assignments, tests, or discussion postings.*** Computers are available on campus in various areas of the buildings and in the Academic Success Center. ***Your computer being down is not an excuse for missing a deadline!!** There are many places to access your class! Our online classes can be accessed from any computer with an internet connection. Contact your instructor immediately if you have computer trouble. If you have technical difficulties in the course, there is also a student helpdesk available to you. The college cannot work directly on student computers due to both liability and resource limitations; however, they can help you get connected to our online services. For help, log in to D2L**.**

Instructor Class Policies

- No food or beverage is allowed in the classroom. This is a college policy.
- Please come to class on time. Take care of personal business prior to class. I do not expect you to leave and return to class (unless there was an emergency, and you explain it to me after class).
- Class time is not for surfing the Web, monitoring Facebook, texting, or catching up on email. You will be asked to leave the class if you continually violate this policy. The same thing applies to cell phone usage for messaging during class.
- Turn off or silence your cell phones and any other electronic devices and put them away. Please, no texting. I think we can all go a little over an hour without contact with the outside world! Leaving class to return calls and coming back is not acceptable. If you have an emergency that requires your cell phone to be on, let me know and we'll work something out.
- Dress appropriately and conduct yourself professionally and with respect toward your peers and the instructor. Please don't talk while the instructor or others are discussing course materials. Participating in the class is the best way to avoid disturbing the class.
- Follow MSU Covid19 behavioral policies and procedures

Change of Schedule

A student dropping a course (but not withdrawing from the University) within the first 12 class days of a regular semester or the first four class days of a summer semester is eligible for a100% refund of applicable tuition and fees. Dates are published in the Schedule of Classes each semester.

Refund and Repayment Policy

A student who withdraws or is administratively withdrawn from Midwestern State University (MSU) may be eligible to receive a refund for all or a portion of the tuition, fees and room/board charges that were paid to MSU for the semester. HOWEVER, if the student received financial aid (federal/state/institutional grants, loans and/or scholarships), all or a portion of the refund may be returned to the financial aid programs. As described below, two formulas (federal and state) exists in determining the amount of the refund. (Examples of each refund calculation will be made available upon request).

Services for Students with Disabilities

In accordance with Section 504 of the Federal Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990, Midwestern State University endeavors to make reasonable accommodations to ensure equal opportunity for qualified persons with disabilities to participate in all educational, social, and recreational programs and activities. After notification of acceptance, students requiring accommodations should make application for such assistance through Disability Support Services, located in the Student Wellness Center, (940) 397-4140. Current documentation of a disability will be required in order to provide appropriate services, and each request will be individually reviewed. For more details, please go to Disability Support Services.

College Policies

Campus Carry Rules/Policies

Refer to: Campus Carry Rules and Policies

Smoking/Tobacco Policy

College policy strictly prohibits the use of tobacco products in any building owned or operated by WATC. Adult students may smoke only in the outside designated-smoking areas at each location.

Alcohol and Drug Policy

To comply with the Drug Free Schools and Communities Act of 1989 and subsequent amendments, students and employees of Midwestern State are informed that strictly enforced policies are in place that prohibit the unlawful possession, use, or distribution of any illicit drugs, including alcohol, on university property or as part of any university-sponsored activity. Students and employees are also subject to all applicable legal sanctions under local, state and federal law for any offenses involving illicit drugs on University property or at University-sponsored activities.

Active Shooter

The safety and security of our campus is the responsibility of everyone in our community. Each of us has an obligation to be prepared to appropriately respond to threats to our campus, such as an active aggressor. Please review the information provided by the MSU Police Department regarding the options and strategies we can all use to stay safe during difficult situations. For more information, visit MSUReady – Active Shooter. Students are encouraged to watch the video entitled "*Run. Hide. Fight.*" which may be electronically accessed via the University police department's webpage: *"Run. Hide. Fight."*

Weather Procedure

In the event of inclement weather, in-person class meetings will be canceled. Alternative assignments, such as online tasks or video-based activities, will be provided to ensure continued learning.

AI-Tool Policy

We encourage students to harness AI tools, like ChatGPT, within the following guidelines:

• **English Writing**: Use AI for grammar and syntax improvement.
• **Drafting & Structuring**: Employ AI to help generate and structure case study drafts.
• **Summarization**: Use AI tools for concise summaries of lengthy case studies.
• **Coding Quality**: Leverage GitHub Copilot (or similar tools) to improve code quality, generate sample codes, and enhance programming efficiency.
However:
• **Original Thought**: While using AI for assistance, students must develop and present their own unique opinions on cases.
• **Academic Integrity**: Understand and support any content from AI tools. Avoid over-reliance and ensure originality. Misrepresentation will face academic consequences.
Leverage AI benefits responsibly and prioritize genuine understanding and original thinking.
Course-specific rules stated in the Exams, Assignments, and Projects sections override the general AI guidelines above.

Grade Appeal Process

Students who wish to appeal a grade should consult the Midwestern State University MSU Catalog

*__Notice:__ Changes in the course syllabus, procedure, assignments, and schedule may be made at the discretion of the instructor.

Course Schedule:

Please keep this syllabus as a reference! Students are responsible for all information in the syllabus and for any changes, which will be announced in class.

Week	Date	Format	Topic
1	01/19/2026 – 01/26/2026	Online	Python Basics I
1	01/21/2026	Lab	Practice: Python Installation, Course Introduction, PyCharm Setup
2	01/26/2026 - 02/02/2026	Online	Python Basics II
2	01/28/2026	Lab	Practice: Writing Functions and Using Loops/Conditionals
3	02/02/2026 - 02/09/2026	Online	Python Collections
3	02/04/2026	Lab	Practice: Manipulating Lists, Tuples, and Dictionaries
4	02/09/2026 - 02/16/2026	Online	Python Modules and Visualization
4	02/11/2026	Lab	Practice: Using math/stats modules, Creating Basic Charts with matplotlib
5	02/18/2026	Lab	Midterm Review
5	02/20/2026 - 02/21/2026	**Exam**	**Midterm Exam: Python Programming**
6	02/23/2026 - 03/02/2026	Online	Data Mining Concepts
6	02/25/2026	Lab	Practice: Exploring a Dataset with Pandas
7	03/02/2026 - 03/09/2026	Online	Data Preparation
7	03/04/2026	Lab	Independent Workday
8	03/09/2026 - 03/16/2026	**Holiday**	**Spring Break**
9	03/16/2026 - 03/23/2026	Online	Data Evaluation
9	03/18/2026	Lab	Practice: Cleaning Data, Encoding Categorical Variables, Normalizing Features
10	03/23/2026 – 03/30/2026	Online	Classification Models I: Regression, Decision Trees, Naïve Bayes
10	03/25/2026	Lab	Practice: Implementing Regression, Decision Trees, Naïve Bayes Using scikit-learn
11	03/30/2026 – 04/06/2026	Online	Classification Models II: SVM, Neural Networks, and Clustering
11	04/01/2026	Lab	Practice: Applying SVM, Neural Networks, and Clustering Techniques
12	04/08/2026	Lab	Midterm Review
12	04/10/2026 – 04/11/2026	**Exam**	**Midterm Exam: Data Mining**
13	04/13/2026 – 04/20/2026	Online	NLP in the Era of Pretrained Models
13	04/15/2026	Lab	Practice: Running Sentiment Analysis with HuggingFace Transformers
14	04/20/2026 – 04/27/2026	Online	Understanding and Preprocessing Business Text
14	04/22/2026	Lab	Practice: Cleaning Text, Tokenizing, and Computing Readability Scores
15	04/27/2026 – 05/04/2026	Online	Applying Pretrained Models to Business Problems
15	04/29/2026	Lab	Practice: Performing Topic Classification and Summarization
16	05/04/2026 – 05/11/2026	Online	NLP Visualization & Evaluation
16	05/06/2026	Lab	Project Showcase & API Integration
Final	05/08/2026 – 05/09/2026	**Exam**	**Final Exam: Natural Language Processing**