OPIM 5671: Data Mining and Time Series Forecasting

Fall 2025

Class:	Wed 18:20 - 21:20 DWTN 134	**Instructor:**	Jaeung Sim jaeung.sim@uconn.edu
Office Hours:	By appointment In person/online	**Prerequisites:**	OPIM 5604 Predictive Modeling

1 Course Overview

This course discusses data mining techniques that can be utilized to analyze large-scale operational data and extract actionable information and knowledge (meaningful patterns, trends, and anomalies) to help make decisions. We will focus on text mining (or natural language processing) and time series forecasting that can be applied to various business domains, such as digital marketing, supply chain management, corporate finance, and asset pricing.

All course materials will be provided through HuskyCT and Github (https://github.com/jaeungs/uconn-opim-5671), and there is no required textbook. Students should be able to implement Python via Google Colab in the classroom, and the details will be introduced in the first class. Note that this course will assume that students have concrete prior knowledge of basic statistics and predictive analysis but not technical familiarity with Python, so this course will introduce the basics and necessary details of Python programming throughout the semester.

The details of the tentative course schedule and content are described in **Table 1**.

2 Content

2.1 Course Overview and Python Basics

In the first week, the course overview and logistics will be presented. A few live quizzes and survey sessions will be conducted to assess students' familiarity with text mining, time series, and machine learning methods. The second week will cover the basics of Python programming, including setting up Google Colab and introducing essential functions and libraries. This foundation will ensure that students are well-prepared to implement text mining and time series analysis—the core focus areas of the course.

2.2 Theme 1: Text Mining and Machine Learning

2.2.1 Text Mining and Singular Value Decomposition

This section introduces the foundations of text mining by focusing on how unstructured text data can be transformed into structured and quantifiable formats for analysis. Students will explore core techniques such as tokenization, vectorization (e.g., TF-IDF), and dimensionality reduction using singular value decomposition (SVD), which serves as a basis for uncovering latent structures in large text corpora. Applications may include document clustering, semantic analysis, and topic modeling. Through hands-on exercises, students will learn how to implement text parsing, filtering, and SVD to extract informative features from text. The focus will be both theoretical and practical, with real-world datasets used throughout.

2.2.2 Predictive Analysis and Feature Generation

This section focuses on generating features from text data for use in predictive modeling. First, it will introduce the essential concepts and metrics to assess predictive models, such as the bias-variance tradeoff,

model interpretability, information criteria, confusion matrix, and receiver operating characteristic (ROC) curve. Building on the conceptual understanding, students will develop a predictive model by extracting features using both supervised and unsupervised methods from text data.

2.2.3 Deep Learning and Text Mining

In this section, students will learn the basic structure of artificial neural networks, recurrent neural networks (RNNs), and and transformer-based models, which have been widely applied to analyze text data. The focus will be on understanding how these models learn from large text corpora and their advantages and disadvantages. The Python hands-on will focus on the details of RNN-based models to build a binary-classification model with text features.

2.2.4 Advanced Topics in Machine Learning

It will primarily discuss the ethical challenges in machine learning algorithms, such as fairness and privacy issues. Regarding the fairness issue, students will learn the sources and consequences of data collection bias discovered by academic research. For instance, they will have in-depth discussions on real-world business cases and field experiments, encompassing both their causes and remedies. With regard to the data privacy, students will learn the recent discussions on right to be forgotten, the background of regulations (e.g., GDPR and CCPA), and the limitations of the database management. Then, they will learn and discuss machine unlearning, a state-of-the-art computational approach to remove the individual data's influence on the model.

2.3 Theme 2: Time Series and Panel Data Analysis

2.3.1 Introduction to Time Series and ARIMA Models

This section introduces students to time series analysis with a focus on identifying temporal patterns and dependencies in sequential data. Core concepts such as stationarity, seasonality, and autocorrelation will be covered, along with basic forecasting techniques. The ARIMA model will serve as the main tool for analyzing and forecasting time-dependent variables. Students will learn to implement and diagnose ARIMA models using real datasets, and develop an intuition for model selection and evaluation in a time series context.

2.3.2 Mathematical Background of ARIMA Models

In a separate week, students will delve deeper into the characteristics of ARIMA models. They will learn mathematical representations and proofs of AR(n) processes, MA(n) processes, and differencing nonstationary time series.

2.3.3 Time Series Regressors and Events

This section extends time series analysis to include explanatory variables and external events, such as policy changes or marketing campaigns. Students will learn how to incorporate covariates into time series models and analyze intervention effects using methods like dynamic regression and transfer function models. Real-world examples will help students understand how to estimate the impact of exogenous shocks on time-dependent outcomes. These skills are particularly useful for evaluating the effectiveness of strategies over time in business, economics, and policy settings.

2.3.4 Causal Inference with Panel Data Analysis

The final module covers causal inference using panel data, where both cross-sectional and time series dimensions are present. Students will be introduced to fixed effects models, difference-in-differences (DiD),

regression discontinuity design (RDD), and synthetic control methods. Emphasis will be placed on identifying causal relationships, testing assumptions, and interpreting results correctly. This section equips students with the tools to answer policy-relevant questions and draw meaningful conclusions from longitudinal data involving firms, regions, or individuals.

3 Grading Schemes

The grading items and their points are summarized in **Table 2**. Overall, class participation has 10% of significance, hands-on-based assignments and projects have 50% of weights, and two exams have 40% of weights in your grades.

3.1 Class Participation (100 points, 10%)

The instructor will call your name in each class. If you fail to respond during the class, it will affect your class participation score. Except for the first class, all students are expected to be present in the classroom. If students leave the classroom without the instructor's consent, their attendance might not be counted.

Virtual attendance is considered in exceptional cases, including school-approved requests for students, extreme events (e.g., hazardous weather conditions, the spread of fatal contagious diseases), and when the instructor cannot provide lectures in person due to valid reasons.

Students have **two tokens** that can compensate for absence. Considering the significance of missing two in-person classes, this course does not accept excuses for further absences. It is fully your responsibility to manage the risks of missing classes more than twice during the entire semester.

3.2 Hands-on Assignments (300 points, 30%)

The individual assignments aim to practice Python programming skills and enhance your understanding of natural language processing and time series analysis. There will be three assignments that require you to handle real-world datasets and apply the concepts and techniques you learn in classes as follows:

- **Hands-on Homework #1 (Predictive Analysis with Text Mining):** In this project, you're going to predict outcomes by applying text-mining techniques to a given dataset. To extract features, you will implement both supervised and unsupervised techniques.
- **Hands-on Homework #2 (Forecasting Stock Prices):** This is a warmup for **Term Project #2**. You'll forecast stock prices of a few firms listed in NASDAQ using various time series forecasting techniques and compare the performances of different model choices.

The detailed instruction will be provided later.

In your assignment, you might get help from digital resources, including online forums (e.g., Stack Overflow), generative AI tools (e.g., ChatGPT), and other online documents. However, you must do your assignment by yourself, and you are not allowed to let other people do your homework instead.

Table 1: Tentative Course Schedule

Date	Format	Topics	Notes
08/27/2025	In person	Course Overview	
09/03/2025	In person	Basics of Python Programming	
		Text Mining and Singular Value Decomposition	
09/10/2025	In person	Basics of Predictive Analysis	
		Supervised and Unsupervised Feature Generation	HW#1
09/17/2025	In person	Deep Learning and Text Mining	
09/24/2025	In person	Recent Challenges in Machine Learning	
10/01/2025	In person	**Term Project Presentation #1**	
10/08/2025	In person	**Mid-term Exam**	
10/15/2025	In person	Introduction to Time Series	
		Exponential Smoothing and ARIMA Models	HW#2
10/22/2025	In person	Mathematical Background of ARIMA Models	
10/29/2024	In person	Time Series with Regressors and Events	
11/05/2024	In person	Causal Inference with Panel Data Analysis	
11/12/2024	In person	**Term Project Presentation #2**	
11/19/2024	In person	Course Review & Final Exam Preview	
11/26/2024	N/A	*Thanksgiving Recess*	
12/03/2024	In person	**Final Exam**	

3.3 Term Project (200 points, 20%)

There will be two term projects, and each project will be done with different teammates. Your projects will be evaluated based on not only your technical excellence but also your communication with the audience. Here is a brief introduction of the projects:

- **Project #1 (Reading Massive Texts with Machine):** You're going to explore real-world text data and draw novel insights using natural language processing. The performance of this project will be evaluated based on the communication about the significance of your problem and insights, as well as the implementation of text mining techniques.
- **Project #2 (Time Series Forecasting Challenge):** You will forecast prices of stocks listed in NASDAQ (**Hands-on Homework #2** will be helpful for this project). In doing so, you might submit two forecasting results, one of which is fully based on time series forecasting, while the other is predicted prices adjusted by your intuition. Your performance will be evaluated based on the results that show the best performance among them.

3.4 Mid-term and Final Exams (400 points in total, 40%)

The exam format and coverage will be announced at least two weeks before the exam date. There will be no makeup exam unless a student is in an extremely exceptional, urgent, and verifiable circumstance. Importantly, the exams aim to test both your programming skills and data science knowledge, expecting the students to digest the overall content in this course.

Table 2: Grading Scheme Overview

Items	Points	Weights
Class Participation	100	10%
Hands-on Assignments	200	20%
Term Projects	300	30%
Mid-term Exam	200	20%
Final Exam	200	20%
Total	1,000	100%

4 Communication Policy

In all types of communication, please address me politely, use appropriate language, and be concise and specific in your questions or requests. Maintaining professionalism and mutual respect helps create a positive and collaborative learning environment. I encourage you to actively listen, thoughtfully engage in discussions, and approach feedback constructively. Together, we can ensure effective communication that supports your academic success.

4.1 Email

I will try my best to respond to your emails typically within 24 hours, except on weekends and national holidays. In your email, please briefly describe who you are, which course you are taking, and why you are reaching out to me. If you need urgent requests or don't get a response from me over 24 hours, I strongly encourage you to send me a reminder.

4.2 Feedback on Assignments

Before grading, I will answer clarification questions only. That is, I will not provide feedback directly related to your scores on your work, such as writing, programming codes, and slides, until I complete grading the assignment.

4.3 Request for Reevaluation

After you receive scores from assignments, quizzes, and the final exam, you might send me a rebuttal letter in one week. In this letter, you should clearly state for which parts and for which reasons you think your scores can change. For class participation and term projects, the instructor's evaluation criteria should be strictly respected, and the scores in these parts will change only if the instructor makes an apparent mistake in the evaluation.

5 Academic Integrity, Responsibilities, and Rights

The University of Connecticut is committed to fostering an intellectual community in which the highest ethical standards of academic, scholarly, and professional integrity prevail. For the detailed background and policies, please refer to the Office of University Compliance.

5.1 University Policies

5.1.1 Definitions

Academic Integrity: a commitment by the University Community to uphold just and ethical behaviors, which includes truthfulness, fairness, and respect (ICAI, 2021).

Scholarly Integrity: a commitment by the University community to both "... research integrity and the ethical understanding and skill required of researchers/scholars in domestic, international, and multicultural contexts. It is also intended to address ethical aspects of scholarship that influence the next generation of researchers as teachers, mentors, supervisors, and successful stewards of grant funds." (p. xix, Council of Graduate Schools, 2012).

Professional Integrity: standards of behavior defined by the various professions in which students are prepared through their degree or certificate programs.

Academic, Scholarly, and Professional Integrity Misconduct: unethical academic and scholarly behavior during a course (e.g., on an assignment or exam), as part of other degree requirements (e.g., requirements regarding placement, capstone or comprehensive exams, or placement exams), or at other times during undergraduate, graduate, or professional study and performance, including during engagement in fieldwork, clinical placements, or research.

These behaviors include **cheating**, **plagiarizing**, **misrepresenting**, **noncompliance**, and so on. Further details are available at the Office of University Compliance.

5.1.2 Policy Statement

All members of the university community, including administrators, faculty, staff, and students, have a shared responsibility to uphold the highest ethical standards of academic, scholarly, and professional integrity and to report any violations of those standards of which they are aware.

Student Expectations: To uphold the principle of academic and scholarly integrity in all aspects of their intellectual development and engagement at the University, students are expected to:

- be responsible for their own work and their own actions related to all academic and scholarly endeavors.
- assume they are to do independent work and seek clarification prior to collaborating with others or using outside resources.
- understand and abide by the standards, protocols, and guidelines to which they must adhere in research, creative, or professional activities.

If students witness or become aware of a violation of academic or scholarly integrity, they are encouraged to communicate this to the appropriate university representative (e.g., faculty, staff, advisor).

A cumulative record is maintained of all academic or scholarly integrity violations and such record will be reviewed and considered as part of subsequent incidences. Individuals engaged in research are expected to follow all standards, rules and regulations that guide the proper conduct of research or creative activity.

5.1.3 Enforcement

Violations of this policy and its related procedures may result in appropriate disciplinary measures in accordance with University By-Laws, General Rules of Conduct for All University Employees, applicable collective bargaining agreements, and the University of Connecticut Student Code.

Note that student misconduct is governed by the University's Student Code, which is administered under the direction of the Division of Student Affairs, as follows:

> *"We never educate directly, but indirectly by means of the environment. Whether we permit chance environments to do the work, or whether we design environments for the purpose makes a great difference." (John Dewey 1933, p. 22).*

Enforcement of its provisions is the responsibility of the Director of Community Standards (for undergraduate students), The Graduate School (for graduate students), and the Office of the Vice President for Research (for research misconduct). Identified misconduct will be routed to the appropriate unit.

5.2 Support for Student Rights

Students with Disabilities: The University of Connecticut is committed to protecting the rights of individuals with disabilities and assuring that the learning environment is accessible. If you anticipate or experience physical or academic barriers based on disability or pregnancy, please let me know immediately so that we can discuss options. Students who require accommodations should contact the Center for Students with Disabilities, Wilbur Cross Building Room 204, (860) 486-2020 or csd@uconn.edu.

Policy Against Discrimination, Harassment, and Related Interpersonal Violence: Discrimination and discriminatory harassment based on any protected class(es), and sexual harassment are prohibited under the University's Policy Against Discrimination, Harassment and Related Interpersonal Violence as follows:

> *"The University of Connecticut (the "University") is committed to maintaining a safe and non-discriminatory learning, living, and working environment for all members of the University community – students, employees, and visitors. Academic and professional excellence can exist only when each member of our community is assured an atmosphere of safety and mutual respect. All members of the University community are responsible for the maintenance of an environment in which people are free to learn and work without fear of discrimination, discriminatory harassment or interpersonal violence. Discrimination diminishes individual dignity and impedes equal employment and educational opportunities."*

For further details, please refer to the Office of Institutional Equity (or contact through (860) 486-2943 or equity@uconn.edu).

Sexual Assault Reporting Policy: The University of Connecticut is committed to cultivating and maintaining an environment free from all forms of sex-based discrimination, harassment, and sexual misconduct. It is important for individuals who are impacted by sexual harassment or any other form of sexual violence to know that the resources contained on this website can offer assistance regardless of whether any formal administrative or criminal process is initiated. This includes help with medical and counseling services, academic and housing support, referrals to legal and confidential advocacy organizations, and assistance with working, visa and immigration, transportation, financial aid matters and more. More information is available at Title IX at UConn. To get help from the University, please contact the Office of Institutional Equity via (860) 486-2943 or equity@uconn.edu.