

COURSE INFORMATION

- INFO 4670, 3 Credit Hours
- Title: **Data Analysis and Knowledge Discovery**
- Class Day & Time: **Tuesday 6.30 PM to 9.20 PM**
- Class Location: **FRLD346**

Instructor Contact Information

- **Dr. Priyan Malarvizhi Kumar**,
Assistant Professor in Data Science, Department of Information Science, University of North Texas.
- Office: DP E236N (By appointment)
- Phone: 940-565-3201
- Email address: priyan.malarvizhikumar@unt.edu

Teaching Assistant and Tutor

- TBA

Course Pre-requisites, Co-requisites, and/or Other Restrictions

- Pre-requisite: Basic knowledge of Excel, Python programming, some data analysis software, or consent of the instructor

Course Format

The course will be conducted in regular face-to-face class. The course utilizes Canvas, UNT's new learning management system, where all course materials will be accessible, and students will submit assignments through the tools provided on Canvas.

Course Description

Introduces the student to data analysis, data mining, text mining and knowledge discovery principles, concepts, theories and practices. Designed for the aspiring or practicing information professional and covers the basics of working with data from a hands-on and practical perspective. Incorporates lecture, discussion, practice of learned concepts, and readings.

Course Goals, Learning Objectives

- The learner will be an active and engaged participant in discussion forums within the learning community by analyzing, constructing/creating, and evaluating the information presented within the textbook, external readings/resources, student research, and class activities.
- The learner will be able to demonstrate an understanding of the fundamental principles, concepts, theories, and practices of data analysis, data mining, and knowledge discovery and discuss the interplay between them.
- The learner will apply gained knowledge to solve real problems with datasets provided using skills developed in the course. The learner will be able to construct an appropriate bibliography in APA6 from scholarly sources of material for study and research.

Materials

Textbook information (**optional**):

- Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 4th Edition by Jiawei Han (Author), Micheline Kamber (Author), Jian Pei (Author).
Link: https://www.amazon.com/Data-Mining-Concepts-Techniques-Management-dp-0128117605/dp/0128117605/ref=dp_ob_title_bk

Textbook information (**Optional**):

- Introduction to Data Mining (2nd Edition) (What's New in Computer Science) 2nd Edition by Pang-Ning Tan (Author), Michael Steinbach (Author), Vipin Kumar (Author).
Link: https://www.amazon.com/Introduction-Mining-Whats-Computer-Science/dp/0133128903/ref=pd_lpo_2?pd_rd_i=0133128903&psc=1
- Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data 1st Edition by EMC Education Services (Editor). Link: <https://www.amazon.com/Data-Science-Big-Analytics-Discovering/dp/111887613X>

Supplementary materials and/or readings (recommended):

Will be added and constantly updated to canvas!

Teaching Philosophy

The instructor will take a problem-solving approach and work together with students to understand data analysis, text mining, and knowledge discovery. We will learn how to solve practical data collecting, text processing, information extraction, and text mining problems. He will monitor the progress of students and is open to suggestions from students. Students are expected to study 12-15 hours per week, and to submit their assignments on time to achieve satisfactory class performance. Interaction between the student and the instructor is guaranteed and strongly encouraged.

TECHNICAL REQUIREMENTS/ASSISTANCE

UIT Help Desk: <http://www.unt.edu/helpdesk/index.htm>

The University of North Texas provides student technical support in the use of Canvas and supported resources. The student help desk may be reached at:

Email: helpdesk@unt.edu

Phone: 940.565-2324

In-Person: Sage Hall, Room 330

Hours are:

- Monday-Thursday 8am-midnight
- Friday 8am-8pm
- Saturday 9am-5p
- Sunday 8am-midnight
- Canvas technical requirements: <https://clear.unt.edu/supported-technologies/canvas/requirements>
- Other related hardware or software necessary for the course: such as headset/microphone for synchronous chats, word processor, etc.

Minimum Technical Skills Needed

Using the Internet and the learning management system Canvas, using email with attachments, creating and submitting files in commonly used word processing program formats, downloading and installing software, using python programs.

Student Academic Support Services

- **Code of Student Conduct:** provides Code of Student Conduct along with other useful links
 - **Office of Disability Access:** exists to prevent discrimination based on disability and to help students reach a higher level of independence
 - **Counseling and Testing Services:** provides counseling services to the UNT community, as well as testing services; such as admissions testing, computer-based testing, career testing, and other tests
 - **UNT Libraries**
 - **UNT Learning Center:** provides a variety of services, including tutoring, to enhance the student academic experience
 - **UNT Writing Center:** offers free writing tutoring to all UNT students, undergraduate, and graduate, including online tutoring
 - **Succeed at UNT:** information regarding how to be a successful student at UNT
- ASSESSMENT & GRADING

Assessments

A student's grade is composed of the following:

- Quizzes (20%)
- Assignments (40%)
- Term Project (40%)

Grading

Class Attendance and Participation. No attendance is required.

Quizzes (20%). There will be **Two quizzes** for this course. Each quiz includes 20 multiple-choice questions, each worth 0.5 points. Students need to complete the quiz in 30 minutes; **students have two attempts for each quiz.**

Assignments (40%). The class will have **Four assignments**. The assignments are designed to help students understand important concepts and gain hands-on data analysis and mining experience using Rapidminer, Weka, Excel, Python, or any other data analysis tools. Assignments should be typewritten, and diagrams should be drawn using graphics software packages such as PowerPoints and Excel, code should be written on Google Colab or Jupiter Notebook.

Term Project (40%). **Report One 20% (submission + presentation); Final Report and the System/programs 20% (submission + presentation + peer evaluation).** The purpose of the term project is to apply what has been taught in this course to process real-world data or information. Students will work in teams (no more than 3 people) to tackle one particular problem assigned or approved by the instructor in Data Analysis, Natural Language Processing, Information Visualization, Machine Learning, Data Mining, and/or Knowledge Discovery. Term project topics and their specifications will be distributed and discussed in class.

Total Points Possible for Semester/Grading Scale = 100

<i>100-90 = A</i>	<i>89-80 = B</i>
<i>79-70 = C</i>	<i>69-60 = D</i>
<i>59 and below = F</i>	

Grading Table

<i>Assignment</i>	<i>Points Possible</i>	<i>Percentage of Final Grade</i>
-------------------	------------------------	----------------------------------

Quiz • 2 quizzes @ 10 points ea.	20 points	20%
Assignment • 4 assignments @ 10 points ea.	40 points	40%
Term Project • Report one and presentation @ 20 points. • Final report presentation and evaluation @ 20 points.	40 points	40%
Total Points Possible	100 points	100%

Tentative Schedule (Subject to Change)

The contents of the course are organized into 16 weeks. Please refer to Table 1 for lessons, topics, and materials.

Table 1. Lessons and Readings

Lessons	Topics
Lesson 1	Course Orientation and Overview Introduction to data analysis and knowledge discovery
Lesson 2	Collecting multivariate heterogeneous data
Lesson 3	Big data and data quality
Lesson 4	Data preprocessing, cleaning, and operation
Lesson 5	Data modeling and database (Data structure)
Lesson 6	Mining frequent patterns, associations, and correlations
Lesson 7	Time series data analysis
Lesson 8	Data mining techniques: regression
Lesson 9	Data mining techniques: clustering
Lesson 10	Data mining techniques: classification
Lesson 11	Data mining and knowledge discovery in practice (applications)

COURSE EVALUATION

Student Evaluation Administration Dates

Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 7, 8 and 9 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via IASystem Notification" (**no-reply@iasystem.org**) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the SPOT website at <http://spot.unt.edu/> or email spot@unt.edu.

COURSE POLICIES

Assignment Policy

Students should submit the assignments and term project reports in canvas.unt.edu: doc (or .docx) files with the code link on GitHub included in the file, also with the code uploaded on GitHub, details will be included in each assignment.

Examination Policy

There are no exams for this course.

Instructor Responsibilities and Feedback

- Helping students grow and learn
- Providing clear instructions for projects and assessments
- Answering questions about assignments
 - Identifying additional resources as necessary
 - Providing grading rubrics
 - Reviewing and updating course content
 - The instructor and TA will respond to students' emails and questions posted to the discussion boards within two days except for the weekends
 - Assignments grades and feedbacks will be returned to the students within one week after the submission deadline.

Late Work and Missed Work

Students are expected to submit assignments and projects on time. If an extenuating circumstance such as a medically diagnosed illness or a family emergency arises, which prevents you from submitting your assignments, you should contact the instructor and the TA as soon as possible before the due date. Late work without the permission of the instructor will receive a grade with a 10% penalty (or 1 point out of 10) per day after the due date. A student who is having trouble with assignments is strongly encouraged to contact the instructor and the TA as early as possible for personal advising.

Course Incomplete Grade

The UNT Graduate Catalog (<http://catalog.unt.edu/index.php?catoid=16>) describes and explains grading policies. A grade of Incomplete (I) will be given only for a justifiable reason and only if the student is passing the course. The student is responsible for meeting with the instructor to request an incomplete and discuss requirements for completing the course. If an incomplete is not removed within the time frame agreed to by the instructor and student, the instructor may assign a grade of F.

Withdrawal

The UNT Graduate Catalog (<http://catalog.unt.edu/index.php?catoid=16>) describes and explains withdrawal policies and deadlines. The UNT semester course schedule lists specific deadlines regarding withdrawal. A grade of Withdraw (W) or Withdraw-Failing (WF) will be given depending on a student's attendance record and grade earned. Please note that a student who simply stops attending class and does not file a withdrawal form may receive an F.

Attendance Policy

Students are encouraged to attend each class meeting. But attendance is not mandatory. Students who attend the session can clearly understand the concepts with practical exposure.

COVID-19 Impact on Attendance

While attendance is expected as outlined above, it is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about COVID-19. Please contact me if you are unable to attend class because you are ill, or unable to attend class due to a related issue regarding COVID-19. It is important that you communicate with me prior to being absent so I may make a decision about accommodating your request to be excused from class.

If you are experiencing any symptoms of COVID-19 (<https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>) please seek medical attention from the Student Health and Wellness Center (940-565-2333 or askSHWC@unt.edu) or your health care provider PRIOR to coming to campus. UNT also requires you to contact the UNT COVID Hotline at 844-366-5892 or COVID@unt.edu for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, is more important.

Students' Responsibility for Their Learning

The students are required to follow course schedule and finish the readings, assignments, quizzes, and term projects. Students are expected to study 12-15 hours per week to achieve satisfactory class performance. Students do not have programming experience are required to find extra materials to study.

UNT POLICIES

Academic Integrity Policy

Academic Integrity Standards and Consequences. According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

ADA Policy

UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one's specific course needs. Students may request accommodations at any time; however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website at disability.unt.edu.

Emergency Notification & Procedures

UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a university closure, please refer to Blackboard for contingency plans for covering course materials.

Retention of Student Records

Student records pertaining to this course are maintained in a secure location by the instructor of record. All records such as exams, answer sheets (with keys), and written papers submitted during the duration of the course are kept for at least one calendar year after course completion. Course work completed via the Blackboard online system, including grading information and comments, is also stored in a safe electronic environment for one year. Students have the right to view their individual records; however, information about student's records will not be divulged to other individuals without proper written consent. Students are encouraged to review the Public Information Policy and the Family Educational Rights and Privacy Act (FERPA) laws and the University's policy. See UNT Policy 10.10, Records Management and Retention for additional information.

Acceptable Student Behavior

Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at **deanofstudents.unt.edu/conduct**.

Access to Information - Eagle Connect

Students' access point for business and academic services at UNT is located at: **my.unt.edu**. All official communication from the University will be delivered to a student's Eagle Connect account. For more information, please visit the website that explains Eagle Connect and how to forward e-mail: **eagleconnect.unt.edu/**.

Sexual Assault Prevention

UNT is committed to providing a safe learning environment free of all forms of sexual misconduct, including sexual harassment sexual assault, domestic violence, dating violence, and stalking. Federal laws (Title IX and the Violence Against Women Act) and UNT policies prohibit discrimination on the basis of sex, and therefore prohibit sexual misconduct. If you or someone you know is experiencing sexual harassment, relationship violence, stalking, and/or sexual assault, there are campus resources available to provide support and assistance. UNT's Survivor Advocates can assist a student who has been impacted by violence by filing protective orders, completing crime victim's compensation applications, contacting professors for absences related to an assault, working with housing to facilitate a room change where appropriate, and connecting students to other resources available both on and off campus. The Survivor Advocates can be reached at **SurvivorAdvocate@unt.edu** or by calling the Dean of Students Office at 940-565- 2648. Additionally, alleged sexual misconduct can be non-confidentially reported to the Title IX Coordinator at **oeo@unt.edu** or at (940) 565 2759.

Important Notice for F-1 Students taking Distance Education Courses

Federal Regulation

To read detailed Immigration and Customs Enforcement regulations for F-1 students taking online courses, please go to the Electronic Code of Federal Regulations website at **http://www.ecfr.gov/**. The specific portion concerning distance education courses is located at Title 8 CFR 214.2 Paragraph (f)(6)(i)(G).

The paragraph reads:

(G) For F-1 students enrolled in classes for credit or classroom hours, no more than the equivalent of one class or three credits per session, term, semester, trimester, or quarter may be counted toward the full course of study requirement if the class is taken on-line or through distance education and does not require the student's physical attendance for classes, examination or other purposes integral to completion of the class. An on-line or distance education course is a course that is offered principally through the use of television, audio, or computer transmission including open broadcast, closed circuit, cable, microwave, or satellite, audio conferencing, or computer conferencing. If the F-1 student's course of study is in a language study program, no on-line or distance education classes may be considered to count toward a student's full course of study requirement.

University of North Texas Compliance

To comply with immigration regulations, an F-1 visa holder within the United States may need to engage in an on-campus experiential component for this course. This component (which must be approved in advance by the instructor) can include activities such as taking an on-campus exam, participating in an on-campus lecture or lab activity, or other on-campus experience integral to the completion of this course.

If such an on-campus activity is required, it is the student's responsibility to do the following:

- (1) Submit a written request to the instructor for an on-campus experiential component within one week of the start of the course.
- (2) Ensure that the activity on campus takes place and the instructor documents it in writing with a notice sent to the International Student and Scholar Services Office. ISSS has a form available that you may use for this purpose. Because the decision may have serious immigration consequences, if an F-1 student is unsure about his or her need to participate in an on-campus experiential component for this course, s/he should contact the UNT International Student and Scholar Services Office (telephone 940-565-2195 or email internationaladvising@unt.edu) to get clarification before the one-week deadline.

Student Verification

UNT takes measures to protect the integrity of educational credentials awarded to students enrolled in distance education courses by verifying student identity, protecting student privacy, and notifying students of any special meeting times/locations or additional charges associated with student identity verification in distance education courses.

See **UNT Policy 07-002 Student Identity Verification, Privacy, and Notification and Distance Education Courses**.

Use of Student Work

A student owns the copyright for all work (e.g., software, photographs, reports, presentations, and email postings) he or she creates within a class and the University is not entitled to use any student work without the student's permission unless all of the following criteria are met:

- The work is used only once.
- The work is not used in its entirety.
- The use of the work does not affect any potential profits from the work.
- The student is not identified.
- The work is identified as student work.

If the use of the work does not meet all of the above criteria, then the University office or department using the work must obtain the student's written permission.

Download the UNT System Permission, Waiver and Release Form