MSDA 665 Big Data Analytics for Business

Spring 2023

Instructor:	Dr. Jinhak Kim	**Email:**	jkim23@niu.edu
Office hours:	By appointment	**Office:**	BH 328V

Course Description

In-depth study of the concepts, methods, and tools for Data Science and Big Data Analytics with a focus on business scenarios. This course is designed to introduce students to advanced analytics techniques that can be applied to big data sets. Topics include python programming, principal component analysis (PCA), data imputation, web scraping, image processing, natural language processing (NLP), and mathematical optimization. For most techniques, students learn how to apply the techniques to solve real-world business problems. All the topics are covered using Python programming on Google Colaboratory along with various libraries such as pandas, NumPy, scikit-learn, BeautifulSoup, Keras, TensorFlow, and gurobipy.

Course Goals and Objectives

Upon the completion of this course, students will be able to

- categorize various machine learning techniques based on the data structure and the purpose of the analyses,
- list various machine learning techniques and explain their mechanisms and fundamentals,
- execute each analytic technique with Python and relevant packages in it,
- apply each analytic technique to solving real-world business scenarios and data sets, and
- analyze and interpret the outcome of the analysis in the business context

Textbooks

There are no required textbooks. The learning materials are available in the lecture notes/slides that I post on our course website.

Course Format & Blackboard

This course adopts an asynchronous delivery modality. All the lectures are delivered asynchronously using the lecture videos. All the course materials including lecture videos are posted on our Blackboard course website. The course consists of the course introduction module (Module 0), which is scheduled one week before the official semester begins, and 10 weekly learning modules (Modules 1 - 10). Only past and active modules are available to students. Visit the Blackboard course website and see the course material listed in Course Contents. To help students keep track of course tasks, I send a weekly announcement and a copy of the announcement will be emailed to your NIU email account. I refer to this announcement as a *module kickoff announcement*. The announcement will include some due dates and a list of activities that you are expected to complete during the module and some due dates. Module kickoff announcements will

be archived on our blackboard website. To access them, click the "Announcements" menu in the list of tools on the left side of the course website. There is only one cumulative exam (final exam) in this course at the end of the semester. The exam takes place asynchronously through our Blackboard course website.

Inquiries

Please feel free to email me if you have any questions. You can find my email address on the first page of the syllabus. For prompt response, please include your full name and the course title (or number) in the email, preferably in the title of the email. Please do not send a message using the "Messages" tool on our Blackboard website. In conclusion, please send me an inquiry directly via NIU email in Outlook and include your full name and the course number in the email. I will answer your question within two business days. But, I typically respond to students' questions much quicker. Besides, you can utilize the "Discussions" tool on our Blackboard course website to ask questions. Students can also respond to the inquiries posted in "Discussions." I will try to check discussion boards frequently, but I can guarantee you to provide immediate responses. For urgent questions, use emails.

Grading

The following tables show how your final scores and point grades are determined. Notice that I do not give C– (C "minus") grade to clarify the passing grade.

Quizzes	20 %	200 points
Homework	45 %	450 points
Final Exam	35 %	350 points
Extra Credits	2 %	20 points
Total	102 %	1020 points

A	93 – %	930 – points
A–	90 – 93 %	900 – 929 points
B+	87 – 90 %	870 – 899 points
B	83 – 87 %	830 – 869 points
B–	80 – 83 %	800 – 829 points
C+	77 – 80 %	770 – 799 points
C	70 – 77 %	700 – 769 points
D	60 – 70 %	600 – 699 points
F	0 – 60 %	– 599 points

Course Activities

Any serious academic dishonesty in any course activity will immediately and strictly result in an "F" grade in the course. Such academic dishonesty includes exam/quiz cheating and project plagiarism.

Quizzes: Ten quizzes (Quizzes 0 – 9) will be given throughout the semester. Quizzes correspond to Modules 0 - 9. There is no quiz in Module 10 because of the final exam scheduled in this module. Quiz 0 includes syllabus reconnaissance, which is designed to get you familiarized with the course policies. Quizzes contribute to 20% ($= 10 \times 2\%$) of your total grade, and each quiz is graded out of 20 points. Each quiz consists of various questions (multiple-choice, true/false, or short-answered questions) to test your knowledge on each learning module. See the Blackboard course website to find the quizzes and their due dates. All quizzes take place asynchronously through our Blackboard course website. Due dates are also available in the "Tentative Course Schedule" section of the syllabus or on our course website.

Homework: There are nine weekly homework assignments throughout the semester. Unlike quizzes, homework assignments are designed for students to have hands-on experience. Homework assignments contribute to 45% ($= 9 \times 5\%$) of your total grade, and each homework is graded out of 50 points.

Final Exam: Final exam takes place in the final exam week. The final exam is cumulative. A study guide and a video review session will be provided at least one week before the final exam date. Homework assignment contributes to 35% of your total grade, and the final exam is graded out of 350 points.

Extra credit: Students can earn up to 2% (=20 points) extra credits. An 1% point extra credit (=10 points) will be granted if you complete the midterm survey administered by the instructor. Another 1% point extra credit (=10 points) will be granted if you complete the university course evaluation and upload the screenshot that includes (1) your name, (2) course number/title, and (3) the completion status.

Late Work Policy

Late work will not be accepted. Illness or other excuses need to be verified - all of which will be accepted at the instructor's discretion. The supporting documentation must be formal and signed by authorized personnel with his/her direct contact. I strongly suggest that you start your quiz and homework as soon as possible. Requesting an extension at the last minute is not accepted in general. For example, a bad internet connection or an illness that occurred a few hours before the due date/time is not considered an acceptable excuse. Start your work as soon as possible to avoid these issues. Students with an unexpected distraction (including illness) must notify the instructor immediately to be considered as an excuse. If such a distraction is too serious to write an email to the instructor, a family member can do that on behalf of the student.

Virtual Office Hours

Students can request office hours by setting up an appointment via email. When an appointment is set up, we will virtually meet using in Zoom, Team, or Blackboard Collaborate. If you scheduled an extra office hour, please be punctual. I may leave the Course Room when the student is late for more than 5 minutes. I reserve the right to refuse an office hour request from a student who late repeatedly. Such students can still use emails to ask questions.

Tentative Course Schedule

The following schedule is subject to change.

Module 0 (Jan 1 - Jan 9): Introduction to MSDA 665

Topics	Activities
Course overview	Quiz 0 (graded) (Due 4 pm, Jan 9)
Data Analytics Overview	Homework Demo (Due 4 pm, Jan 9)
Software Setup	(mandatory, ungraded)

Module 1 (Jan 9 - 16): Excel and Python Preliminaries

Topics	Activities
Excel: Functions, fill handle	Quiz 1 (Due 4 pm, Jan 16)
Excel: Absolute referencing	Homework 1 (Due 4 pm, Jan 16)
Python: Variables, list, array, range	
Python: Dictionary, data frame	
Python: if/for/while statements	
Python: User-defined function, data import	

Module 2 (Jan 16 - 23): Principal Component Analysis

Topics	Activities
Measuring variability	Quiz 2 (Due 4 pm, Jan 23)
Introduction to PCA	Homework 2 (Due 4 pm, Jan 23)
Calculating Principal Components	
PCA as a dimension reduction technique	

Business Scenario: Analysis of U.S. Treasury yield rates

Module 3 (Jan 23 - 30): Data Imputation Part 1

Topics	Activities
Data imputation applications	Quiz 3 (Due 4 pm, Jan 30)
Missing data mechanisms	Homework 3 (Due 4 pm, Jan 30)
Measuring performance	

Business Scenario: Image recovery, movie recommendation system

Module 4 (Jan 30 - Feb 6): Data Imputation Part 2

Topics	Activities
k-fold cross validation	Quiz 4 (Due 4 pm, Feb 6)
Understanding imputation methods	Homework 4 (Due 4 pm, Feb 6)
Image recovery	

Business Scenario: Image recovery, movie recommendation system

Module 5 (Feb 6 - 13): Web Scraping

Topics	Activities
Basics of HTML	Quiz 5 (Due 4 pm, Feb 13)
Navigating on HTML trees	Homework 5 (Due 4 pm, Feb 13)
Parsing HTML documents	

Business Scenario: Scraping historical yield rates and S&P indices

Module 6 (Feb 13 - 20): Deep Learning Part 1

Topics	Activities
Biological Neural Network	Quiz 6 (Due 4 pm, Feb 20)
Activation functions	Homework 6 (Due 4 pm, Feb 20)
Forward feeding & backpropagation	
Simple regression using deep learning	
Importing data and image visualization	
Overfitting	

Business Scenario: Fashion image classification

Module 7 (Feb 20 - 27): Deep Learning Part 2

Topics	Activities
Convolutions and pooling	Quiz 7 (Due 4 pm, Feb 27)
Text Tokenization	Homework 7 (Due 4 pm, Feb 27)
Padding	Course Survey (Due 4 pm, Feb 27)
Embedding	
Building a text classification model	

Business Scenario: IMDB Review & Twitter US Airline sentiment analysis

Module 8 (Feb 27 - Mar 6): Linear/Integer Programming

Topics	Activities
Introduction to mathematical optimization Solving a simple LP with a geometric method Formulating with OpenSolver Double-indexed formulation	Quiz 8 (Due 4 pm, Mar 6) Homework 8 (Due 4 pm, Mar 6)

Business Scenario: Production planning, transportation problem

Module 9 (Mar 6 - 13): Large-Scale Optimization

Topics	Activities
Formulating with Gurobi and Python Declaring variables and constraints	Quiz 9 (Due 4 pm, Mar 13) Homework 9 (Due 4 pm, Mar 13) Course Evaluation (Due 4 pm, Mar 13)

Business Scenario: Delivery routing problem

Module 10 (Mar 13 - 20): Review/Final Exam

Topics

Review Session
Final Exam (due 4 pm, Mar 20): Take the exam on any day in the time period.

Course Civility

1. You are expected to do all the tasks on your own. Cheating, plagiarism, and any other form of academic dishonesty will not be tolerated and will immediately result in an F in this course.
2. Classroom discussion (in a discussion board, a group activity, or an email) should be civilized and respectful to everyone and relevant to the topic we are discussing. Classroom discussion is meant to allow us to hear a variety of viewpoints. This can only happen if we respect each other and our differences.

Respect for Diversity

I intend that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength, and benefit. I intend to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you."

Expectations for respectful dialogue

We all come to the table with differing experiences and viewpoints, which means that we have so much to learn from each other! In order to get the most out of this opportunity, it is important that we don't shy away from differences. Rather, we should show respect for differences by seeking to understand, asking questions, clarifying our understanding, and/or respectfully explaining our own perspective. This way, everybody comes away with new perspectives on the issue and respecting others with different values or beliefs.

Statement of Accessibility

If you need accommodation for this class, please contact the Disability Resource Center as soon as possible. The DRC coordinates accommodations for students with disabilities. It is located in the Campus Life Building, Suite 180, and can be reached at 815-753-1303 or drc@niu.edu.

Also, please contact me privately as soon as possible so we can discuss your accommodations. Please note that you will not be required to disclose your disability, only your accommodations. The sooner you let me know your needs, the sooner I can assist you in achieving your learning goals in this course.

Policy on Incompletes

A final grade of "Incomplete" is given only under unusual extenuating circumstances; the instructor will decide whether to grant an Incomplete, based on the merits of individual requests. In order for the instructor to entertain an incomplete request, the student must have completed 80 percent of the course assignments. Please also note that a grade of Incomplete carries a penalty: it will result in a final grade that is one letter grade lower than otherwise earned.

NIU Academic Integrity Policy

To make authoritative statements without giving credit to the author of those statements is considered plagiarism and violates the academic integrity policy of Northern Illinois University: Good academic work must be based on honesty. The attempt of any student to present as his or her own work that which he or she has not produced is regarded by the faculty and administration as a serious offense. Students are considered to have cheated if they copy the work of another during an examination or turn in a paper or an assignment written, in whole or in part, by someone else. Students are responsible for plagiarism, intentional or not, if they copy material from books, magazines, or other sources without identifying and acknowledging those sources or if they paraphrase ideas from such sources without acknowledging them. Students responsible for, or assisting others in, either cheating or plagiarism on an assignment, quiz, or examination may receive a grade of F for the course involved and may be suspended or dismissed from the university (2020-2021 Undergraduate Catalog, p.69).

Writing Center

Students who would like assistance developing their writing style and/or skills are encouraged to do so through the NIU Writing Center located in Stevenson South Towers - Lower Level; (815)753-6636.

Northern Illinois University: A Community of Learners

The NIU Community of Learners builds on knowledge, practice, and reflection to produce Exemplary Educators. The Community encompasses scholars, education professionals, and pre-service teachers in an interaction that develops the strengths that embody excellence in education. These strengths include creative and critical thinking, scholarship, and caring. Application of these strengths emerges through the collaborative efforts of a diverse community, which supports lifelong learning.